

5-2012

# The effects of accountability on leniency reduction in self- and peer ratings on team-based performance appraisals

Brettney Smith

Clemson University, [brettns@clemson.edu](mailto:brettns@clemson.edu)

Follow this and additional works at: [https://tigerprints.clemson.edu/all\\_dissertations](https://tigerprints.clemson.edu/all_dissertations)



Part of the [Psychology Commons](#)

---

## Recommended Citation

Smith, Brettney, "The effects of accountability on leniency reduction in self- and peer ratings on team-based performance appraisals" (2012). *All Dissertations*. 895.

[https://tigerprints.clemson.edu/all\\_dissertations/895](https://tigerprints.clemson.edu/all_dissertations/895)

This Dissertation is brought to you for free and open access by the Dissertations at TigerPrints. It has been accepted for inclusion in All Dissertations by an authorized administrator of TigerPrints. For more information, please contact [kokeefe@clemson.edu](mailto:kokeefe@clemson.edu).

THE EFFECTS OF ACCOUNTABILITY ON LENIENCY REDUCTION IN SELF-  
AND PEER RATINGS ON TEAM-BASED PERFORMANCE APPRAISALS

---

A Dissertation  
Presented to  
the Graduate School of  
Clemson University

---

In Partial Fulfillment  
of the Requirements for the Degree  
Doctor of Philosophy  
Industrial-Organizational Psychology

---

by  
Brettney DāSean Smith  
May 2012

---

Accepted by:  
Dr. Fred Switzer, III, Committee Chair  
Dr. Mary Anne Taylor  
Harold E. Cheatham, Ph.D.  
Dr. Claudio Cantalupo

## ABSTRACT

The purpose of the present study was to assess the effects of accountability on leniency reduction on self- and peer ratings on team-based performance appraisals when they were used for different purposes (developmental versus evaluative purposes). Accountability was operationalized as participants being told they would have to justify their self- and peer ratings of team behaviors to a local nuclear process control plant supervisor (lab study) or to their professors (field study). In the lab study, purpose was operationalized as participants being told that they would have to complete the Team Behaviors Form (TBF) to receive course credit. In the field study, purpose was operationalized as participants reading (on the TBF) that their ratings would count toward their own and peers' final grade. The results provided partial support for the proposed hypothesis that accountability may help in reducing leniency in team-based performance appraisals and offers evidence for the potential effects of purpose in team performance appraisals. Implications of these results, limitations, and ideas for future research are discussed.

## ACKNOWLEDGMENTS

First, I would like to thank God for helping me to achieve this significant milestone in my life. I would also like to thank my Mother for all of her love and support over the years. Additionally, I would like to thank my advisor, Dr. Fred Switzer, III, for all of his guidance throughout my undergraduate and graduate school career. He has helped me grow from an undergraduate REU student to a working professional in the field. In addition to Dr. Switzer, I appreciate the practical knowledge that Dr. Mary Anne Taylor has given me over the years because it has helped me to apply theory within organizational settings in a realistic way and in a way that makes sense.

Last, but not least, I would thank my additional dissertation committee members, Dr. Claudio Cantalupo, and Harold Cheatham, Ph.D., for their helpful insights and comments which have helped make this manuscript more digestible and comprehensible.

## TABLE OF CONTENTS

	Page
TITLE PAGE .....	i
ABSTRACT.....	ii
ACKNOWLEDGMENTS .....	iii
LIST OF TABLES .....	vi
LIST OF FIGURES .....	vii
CHAPTER	
I.    INTRODUCTION .....	1
Literature Review of Team-Based Performance Appraisals .....	6
Team performance .....	6
Antecedents to team performance appraisals.....	8
Team performance appraisals: Design and implementation .....	12
Assessing reliability and validity of team performance.....	13
Measuring team member effectiveness and team performance...	15
Implementation of team performance appraisals in different environments.....	19
Variables that influence effectiveness of team performance appraisals.....	22
Self and Peer Team-Based Performance Appraisals .....	26
Self-other comparisons .....	26
Peer-other comparisons.....	29
Peer ratings and organizational outcomes.....	31
Purpose.....	33
Purpose and individual performance appraisals .....	34
Purpose and team-based performance appraisals .....	36
Accountability.....	37
Summary: Literature Review .....	52
Team-based performance appraisals: Self- and peer ratings .....	52
Purpose.....	53
Accountability.....	54
Purpose: Current Study .....	54
Hypotheses .....	56

II.	METHOD .....	59
	Overview of Method Section .....	59
	Method .....	59
	Participants.....	59
	Measures .....	60
	Pilot Study.....	62
	Design .....	65
	Study 1: Laboratory Study .....	65
	Apparatus .....	65
	Procedure .....	67
	Accountability and purpose conditions.....	68
	Study 2: Field Study.....	70
	Procedure .....	70
	Team Behaviors Form: Developmental Purpose .....	71
	Team Behaviors Form: Evaluative Purpose .....	73
III.	RESULTS .....	76
	Analyses .....	76
	Lab Study Results .....	78
	Field Study Results .....	80
IV.	DISCUSSION .....	82
	Limitations and Future Research .....	88
	Conclusion .....	95
	APPENDICES .....	97
	A: Protocol: Process Control Plant Simulator .....	98
	B: TBF Dimensions Defined Doc.....	103
	C: Lab TBF: Accountability-Evaluative.....	104
	D: Lab TBF: Accountability-Developmental .....	106
	E: Lab TBF: No Accountability-Evaluative.....	108
	F: Lab TBF: No Accountability-Developmental .....	110
	G: Field TBF: Accountability-Evaluative.....	112
	H: Field TBF: Accountability-Developmental .....	115
	I: Field TBF: No Accountability-Evaluative.....	118
	J: Field TBF: No Accountability-Developmental .....	120
	K: Field Study IRB Approval Form.....	122
	REFERENCES .....	123

## LIST OF TABLES

Table	Page
1 Descriptive statistics for the lab study results.....	78
2 Descriptive statistics for the field study results .....	80

## LIST OF FIGURES

Figure	Page
1 Hypothesized effects of accountability, purpose, and appraisal source on ratings of team behaviors.....	56
2 XPlant Process Control Plant Simulator Diagram .....	67
3 Effects of accountability, purpose, and appraisal source on ratings of team behaviors – Lab Study .....	79
4 Effects of accountability, purpose, and appraisal source on ratings of team behaviors – Field Study .....	81



## CHAPTER ONE

### INTRODUCTION

Since the mid-twentieth century, applied psychologists have been interested in the study of performance appraisals. From the 1950s until the 1980s, the focus of performance appraisals was on the psychometric properties (e.g., reliability) of rating instruments. From this focus emerged several different forms of performance appraisals, such as graphic rating scales and behaviorally anchored rating scales. During the '80s, the focus of appraisal research then shifted to the study of information processing and cognition. This research focused on how raters assign ratings to the target of interest. Some researchers, such as Bretz and colleagues (1992) felt that this intense focus on cognitive processing issues led to a research-practitioner gap. They felt that the research being conducted did not meet the needs of evolving and complex organizations. (Bretz et al., 1992).

Addressing these concerns, recent research has considered the importance of organizational politics and social context in performance appraisals. Rather than focus purely on psychometric properties of performance appraisals, some researchers have examined situational influences, particularly the organizational context and rater goals, that affect employees' ratings of a target and of themselves. Additionally, this line of research has focused on ratees' reactions to performance appraisal feedback and how these reactions influence their subsequent ratings of others' behavior and their own future job performance (Levy & Williams, 2004). Only recently have researchers begun to address these issues within the performance appraisal context.

All of the aforementioned issues have resulted from research on individual-based performance appraisals. Many organizations are now implementing teams to carry out many short- and long-term assignments. These teams can better enable organizations to quickly achieve their mission critical functions. When designing and implementing teams, organizations often do not distinguish between different team categories that are instrumental to effective team performance, including team processes, such as shared mental models (i.e., the extent to which team members are on the same page) and team based outcomes (Scott & Einstein, 2001). Also, these organizations are trying to assess these teams' job performance with appraisal systems that are designed to measure only individual-level job performance. So, there is a need to examine factors that are instrumental to effective team performance and to examine research conducted on team-based performance appraisals.

When creating team-based performance appraisals, it is important to examine issues that pertain to the source of performance appraisals. Supervisor, self-, and peer ratings are often the main sources of individual and team-based performance appraisals. Such appraisals differ based on their forms of measurement, rater identification, and their purpose. Although many organizations believe that supervisor performance ratings are the most objective measures of employee performance, many practitioners and researchers have challenged this belief. Specifically, they have challenged the validity of supervisor ratings because many supervisors do not have the opportunity to directly observe their employees; many supervisors' ratings are flawed because of their difficulty in providing ratings of individual performance for highly interdependent tasks; and

because of the role that the subordinate-supervisor interpersonal relationship can play in these ratings (Fedor, Bettenhausen, & Davis, 1999). Thus, peer and self-ratings can serve as supplements to supervisors' ratings, and they can allow employees to receive pertinent feedback instrumental to improving their performance which in turn could affect administrative decisions, such as those regarding pay raises or promotions.

Peer ratings can enhance individual and team-based performance appraisal systems because they can serve as multiple data points that create a holistic picture of an employee's job performance, a picture that supervisor ratings alone cannot create (Drexler, Beehr, & Stetz, 2001). Due to the increasing reliance on teams and workgroups, organizations are incorporating peer ratings into team-based performance appraisals. In principle, peer ratings are useful because peers work closely with the appraisal target and they are able to diagnose the target's proficiency in job-related behaviors. In fact, it has been argued that peers may be the only ones who can accurately assess employees' job performance due to their daily interactions with and observations of the target employees. (Fedor et al., 1999). In addition to the use of peer ratings, self-ratings can allow for employee input into the appraisal system. When self-ratings are incorporated into individual and team-based performance appraisals, they allow employees to evaluate their own job performance levels on various tasks and job dimensions. When employees self-assess their performance, they can take more responsibility for their job behavior which may cause them to become more committed to enhancing their job performance. So, their input into the appraisal system can help them contribute to the overall mission and goals of the organization. Given the potential that self-ratings can serve as a basis for

administrative decisions, this form of employee input (into the performance appraisal system) can increase perceptions of fairness, levels of job satisfaction, and this input can reduce interpersonal conflict between supervisors and subordinates when the subordinates receive negative feedback. Further, self-appraisals can reduce the general anxiety and tensions that employees have about the performance appraisal process. With these self-ratings, employees have a voice or say about the final ratings that they receive from their supervisor because they are able to dispute these ratings based on their own self-evaluations (Roberts, 2003).

Although there are many positive benefits associated with peer and self-ratings, there are also many reservations about their use in performance appraisal systems. Peer ratings have the potential to create additional work stress and anxiety. Employees may not want to take on this task and it may be an unwanted duty to the many other job tasks they have to complete on a daily basis. If employees are unwilling to take on these additional roles as peer evaluators or if these employees do not want to risk threatening the interpersonal dynamics of their teams or workgroups, they are more likely to provide ratings that are inflated and that are inaccurate assessments of their peers' job performance. When peer ratings are used in teams or workgroups, they have the potential to damage interpersonal relationships within the team, leading to a decrease in workgroup productivity or to the dissemination of the team (Bamberger, Erev, Kimmel, & Oref-Chen, 2005). Also, peer ratings may be seen as an illegitimate source of performance evaluation because they violate the psychological contract between subordinates and their supervisors—an informal belief that it is the supervisor's responsibility to appraise the

subordinate's job performance, and that the results will be used as a primary basis for administrative decisions. This perception can lead to a decrease in employees' perceptions of fairness, a negative perception of the organization, and a decrease in their job performance (Fedor et al, 1999). Similar rating biases can also result from the use of self-ratings in organizational settings. In fact, many organizations are reluctant to use self-ratings because of the common perception that these ratings are prone to leniency bias. Employers believe that their employees will provide inaccurate assessments of their job performance, and research has shown that self-ratings are often incongruent with other sources of performance appraisal ratings (Heidemeier & Moser, 2009).

Based on the foregoing research, many peer and self-ratings for both individual and team-based performance appraisals are not reflections of employees' actual job performance. Researchers are currently investigating ways to get these appraisals to reflect employees' actual job performance. Specifically, researchers are investigating ways to reduce leniency bias in peer and self-performance ratings. Some researchers have found that accountability is a good way to reduce leniency in self-ratings of job performance (Smith & Switzer, 2009). Researchers and laypeople alike overuse the term accountability as the end-all solution to many organizational and societal problems without giving proper explanation as to what it entails. In the field of psychology, accountability is operationalized as the belief that people have to justify their actions to someone else (Lerner & Tetlock, 1999).

Given the numerous beneficial effects that peer ratings, self-ratings, and accountability can have on individual performance appraisals, it is important to examine

these issues in the context of team-based performance appraisals. The purpose of the current literature review is to examine published research on team-based performance appraisals, peer and self-ratings on team-based performance appraisals, appraisal purpose, and to examine research on accountability. The review of these bodies of research provides a basis for discussion about whether accountability can reduce leniency effects in peer and self-ratings on team-based performance appraisals.

### **Literature Review of Team-based Performance Appraisals**

Teamwork has become the core makeup of many organizations, and the use of work groups or work teams spans across a variety of academic and applied contexts ranging from group projects in classrooms, IT customer-oriented group projects, and real-time teams in military settings (Salas, Cooke, & Rosen, 2008). Based on the increased use of teams, organizations are faced with the additional challenge of creating team-based performance appraisals. Although there are very few prescriptions for team-based performance appraisals, many existing appraisals (e.g., individual performance appraisal systems that include peer ratings) are not designed to measure team behaviors and overall performance. To design effective team-based performance appraisals, it is important to review current research on team performance, including dimensions of effective team performance.

**Team Performance.** In order to design effective team-based performance appraisals, one must know what constitutes effective team performance. When defining effective team performance, it is important to separate the assessment of team process variables from the assessment of team effectiveness in terms of results-oriented variables.

Separating these assessments can help researchers and employers provide more meaningful feedback to team members depending on the purpose of the team performance appraisal. If the purpose of the team-based appraisal is for developmental feedback, research has shown that it is best to assess team processes; if the purpose of the appraisal is to make administrative decisions, research has shown that it is best to assess team effectiveness (specifically team member effectiveness).

Researchers have recently examined the core components of team processes. After conducting a review on teams research over the past twenty years, Salas and colleagues (2005) have discovered eight essential components to effective team processes: Team leadership, mutual performance monitoring, backup behavior, adaptability, team orientation, shared mental models, closed loop communication, and mutual trust. Team leadership is defined as the capability of a team member to delegate group tasks to other team members, evaluate the level of team performance, foster the enhancement of team members' knowledge, skills, and abilities, and create a positive environment for the team. Mutual performance monitoring is defined as team members' ability to create an environment in which they share knowledge of the responsibilities and duties of each other to track individual and team performance. Backup behavior is conceptualized as team members' proficient knowledge of each others' job duties. This proficient knowledge enables them to foresee the needs of their team members, such as distributing the workload of a team member or team members who are unable to perform their tasks or job duties. Adaptability is a skill that causes team members to be able to shift their tactics and behaviors in response to factors internal and external to the team.

Team orientation is the extent to which team members value team goals over individual goals and team members' willingness to work collectively as a unit to accomplish these team goals. Shared mental models consist of a common framework in which team members are aware of the relations among team duties and goals and of the interpersonal relationships that team members have with each other. Mutual trust is team members' perceptions of their ability to execute team tasks and safeguard the psychological, social, and emotional interests of each other. Closed-loop communication is a form of communication that involves the direct transaction between the sender and the recipient regardless of the communication modality (Salas, Sims, & Burke, 2005). Regarding team member effectiveness, Loughry and colleagues (2007) conducted exploratory and confirmatory factor analyses on team based data of 153 undergraduate participants working in teams of four to five members and they found five dimensions of team member effectiveness that align with Salas and colleagues' (2005) dimensions of effective team performance: Contributing to the team's work, interacting with teammates, keeping the team on track, expecting quality, and possessing the relevant knowledge, skills, and abilities (KSAs).

**Antecedents to team performance appraisals.** Prior to designing team performance appraisals, it is useful to investigate team member selection because poor team member composition will probably result in poor team performance. The empirical literature on team member selection is scant. Miller (2001) examined team member selection by determining the reliability and validity of Stevens and Campion's (1994) Teamwork Test, a measure supposedly used to select highly qualified individuals into



teams based on supposed team KSAs. The Teamwork Test is a 35 multiple-choice item measure that contains five key team based interpersonal and self-management KSAs: Conflict resolution, collaborative problem-solving, communication; goal setting and performance management; and planning and task coordination. Based on regression analyses of team average and team variance scores of 176 undergraduate participants who made up 44 teams, Miller (2001) found that the measure did not significantly relate to team effectiveness, measured by the students' group project grades. Although some theories and measures exist for team member selection, there are few studies that find empirical support for these theories and measures. So, until empirical support arises, researchers and practitioners should assess team processes to improve overall team performance.

Prior to the design of team performance appraisals, researchers and practitioners must ensure that job tasks, duties, and responsibilities are actually best suited for teams or workgroups. It can be argued that it may not be fruitful to design team performance appraisals for organizations if the job-related tasks or jobs, generally speaking, are not team appropriate. In other words, if the tasks and job duties can be executed by individuals—and not by teams—then employees should be assessed individually on their performance, not as a team. Arthur Jr., Edwards, Bell, Villado, and Bennett Jr. (2005) investigated the validity of three team tasks analysis scales that were designed to assess whether a team was required to complete group of tasks or a job duty. These scales assessed team-relatedness of these job tasks and the amount of team workflow. Over a two week timeframe, 52 males working in 4-person teams rated, on based on a range

from 0 to 100%, the team-relatedness and the perceived workflow of a group of tasks based on a computer simulated combat mission. Based on interrater reliability indices of these scales, the authors found support for these scales of team-relatedness of tasks and team workflow pattern from tasks were successful in determining whether a group of tasks and jobs were team-appropriate. Additionally, the authors found that team workflow ratings based on jobs rather than a group of tasks were significantly related to team performance (Arthur, Jr. et al., 2005).

If organizations want to successfully implement team performance appraisals, they must anticipate employees' potential reactions to team performance appraisals. Organizational researchers have examined employee preferences regarding these appraisals. Waldman (1997) conducted two studies on potential predictors of employee preferences for group-based performance appraisals based on employees' achievement orientation (high versus low), the level of collectivistic work norms within organizations, and the nature of employees' work design (i.e., teamwork design versus individual-based design). Based on regression analyses of 276 employees from Canadian utility and transportation companies, and on regression analyses of employees from a large department of the Canadian federal government, the author found that group performance appraisals were preferred over 360-degree performance appraisals when the work was designed as a team-based effort (Waldman, 1997).

Researchers have also examined preferences for team based appraisals in academic settings. Hoffman and Rogelberg (2001) wanted to determine college students' preferred project group grading procedures. They assigned 360 undergraduate

participants to view 1 of 12 versions of a hypothetical syllabus for a college course based on the students' grade point averages (high versus low), the percentage that the group work accounted for the students' grades, and the opportunity to provide input into their group grades, and the evaluation target (individual versus group performance). After conducting a multiple analysis of variance, the authors found that students with low GPAs were more likely to enroll in a course that contained group-based evaluations when these evaluations accounted for a significant portion (50%) of their final grades.

Generally speaking, students were more likely to enroll in a hypothetical course when both individual and group performance evaluations were used in the grading procedures (Hoffman & Rogelberg, 2001). Barfield (2002) examined college students' preferences for group evaluations based on employment status (part- versus full-time) and based on different student age groups: older students (28-47 years old), middle students (23-27 years old), and younger students (18-22 years old). From an analysis of variance on data from 230 undergraduate participants, the author found that older students were more likely to be dissatisfied with group performance evaluations in comparison to younger students and that older part-time students were the most dissatisfied of all students concerning their overall group performance appraisal outcomes (Barfield, 2002). Thus based on Barfield's finding, educators might be well advised to take student age and employment status into consideration when assigning students to teams and when designing group or team-based performance appraisals. Older students may believe they are more knowledgeable and can execute assignments on teams better than their younger

teammates and will be dissatisfied if their overall group evaluations do not reflect these perceptions.

**Team performance appraisals: Design and implementation.** Some researchers have attempted to provide prescriptions for the creation and implementation of effective team performance appraisals. Several researchers have attempted to provide prescriptions for the creation and implementation of team performance appraisals. Zigon (1994) believes that there are five keys to the successful design and implementation of team performance appraisals:

1. Matching team outcomes with the organization's goals.
2. Making the team appraisal customer oriented and assessing the team's performance on customer satisfaction.
3. Assessing employee performance at the team and individual level.
4. Aiming for verification in appraisal ratings.
5. Training the team to develop its own appraisals.

In order to develop team performance appraisals Zigon (1995) believed that practitioners must first create team performance standards following a seven-step process in which they 1) examine previous performance appraisals, 2) specify the assessment points for the teams, 3) note team members' successful contributions to the team, 4) have the team evaluate these contributions in terms of their contribution, 5) create individual and team performance appraisals and 6) standards, and 7) come up with a way to monitor team performance. Simonds Jr. and Bell (1997) carried out Zigon's (1995) principles for creating performance standards for the PECO Energy company. They created a new

appraisal system that was results-oriented and that reflected expectations for employees' behavioral-based job performance. Although the researchers offered advice on how to analyze successful employee job performance, they failed to report whether their new appraisal system was successful and they failed to offer any advice on how practitioners should ensure that the appraisal system aligns with the organizational culture, mission, and goals. A pure results-oriented team performance appraisal may exhibit criterion deficiency because it may not capture the holistic performance of the team; a results-oriented team performance appraisal may be inappropriate if the so-called objective criterion measures can be influenced by external factors beyond the control of team members.

**Assessing reliability and validity of team performance.** In relation to the measurement of the validity and reliability of team performance appraisals, Valle and Davis (1999) attempted to use particular metrics to increase the reliability and validity of a team-based peer performance appraisal system. Forty-four undergraduate students in seven student teams completed individual exams at the beginning of class and then completed a team performance based measure. The participants rated each other using a peer behaviorally anchored rating instrument that measured student contributions to the team's work, attendance, test preparation, interpersonal skills, and overall performance. Based on intraclass correlations between student raters within teams and based on the relation between the peer evaluations and individuals' test scores, the authors found support for the reliability and validity of this team performance appraisal. Yet, based on Arthur et al.'s (2005) study concerning the appropriateness of groups of tasks, it seems

that these authors should have determined whether these testing tasks were team-appropriate. One could argue that these tasks, regardless whether they are taken individually or as a group, are autonomous and hence the utility of the team-based peer appraisal is minimized.

In addition to measuring the reliability and validity of team-based peer appraisals, some researchers have attempted to apply best practices in performance appraisal systems to the measurement of team mental models (TMMs) (Webber, Chen, Payne, Marsh, & Zaccaro, 2000). Webber et al. (2000) investigated whether the indices of accuracy and similarity, particularly regarding TMM measurement would predict future team performance. The authors used content validity procedures to develop the measure because they used subject matter experts (SMEs) who were tenured high school coaches with at least 15 years of experience to develop the TMM questionnaire. After its development, 147 members of 24 community basketball teams completed the measure. TMM similarity was measured by coefficient alpha and the interrater agreement,  $rwg(j)$ , statistic. TMM accuracy was measured using one index which was the difference between the team members' scores on individual items and the ratings assigned by the SMEs. Based on hierarchical regression analyses, the authors found that the interrater agreement index was marginally significant in its relation to the teams' subsequent performance. Neither the coefficient alpha nor the accuracy measure related to team performance (Webber et al., 2000). Although the examination of team process variables is fruitful, this study shows that it is inappropriate to assess only team processes variables

when attempting to determine team performance; team member effectiveness variables should be used instead.

Brannick, Prince, Prince, and Salas (1995) attempted to evaluate the construct validity of measures of team process. Fifty one aircrew teams were evaluated by independent judges (undergraduate students, graduate students, and authors) on 6 process variables (assertiveness, decision making, situational awareness, leadership, and communication). The authors used a multitrait-multimethod (MTMM) analysis to evaluate the construct validity of the team process appraisal, and they treated the judges and the experimental scenarios (simulated pilot exercises) as method variables. The MTMM analysis showed that there was an acceptable amount of convergent and discriminant validity for the team process appraisal when the judges were treated as a method variable in comparison to when the experimental scenarios were treated as method variables. The results suggested that the team process appraisal was sensitive to team composition characteristics, and the results show the importance of the use of multiple observations to assess team process performance (Brannick et al., 1995).

**Measuring team member effectiveness and team performance.** In comparison to studying the design and implementation of team process performance appraisals, researchers have also attempted to study the design and implementation of team member effectiveness. Loughry and colleagues (2007) appear to be the only authors to develop and test a theoretically based measurement of team member effectiveness. The measurement is termed the Comprehensive Assessment of Team Member Effectiveness (CATME) which has 87 items that assess 29 types of contributions by team members (3

items for each of the 29 types). The authors gathered these items based on their literature review of teamwork which entailed research on team effectiveness, peer evaluations, and the lack of acceptable peer evaluation measures. In order to examine these items, the authors used survey data from two large samples of undergraduate students to determine the importance of the particular items as they pertained to an effective team member. Based on exploratory and confirmatory factor analyses of the survey data, the authors found that they could be placed within five dimensions: teamwork contribution, teammate interaction, keeping the team on track, expecting quality, and having relevant KSAs.

Zhang and Ohland (2009) wanted to determine best practices in assigning individual scores on a group project by determining the validity of group grading procedures: within group adjustment, partial adjustment, between group adjustment, and the expected contribution methods. Using a Monte Carlo simulation and a student sample of teams, hypothetical and real participants assessed their own and others' ratings on particular group projects. The authors found that using the CATME as the instrument for self and peer appraisals to adjust team members' contributions to the group project was an effective way to enhance the validity of group grades, and they found that using between and within-group adjustments was the most effective way to increase the validity of group grading procedures (Zhang & Ohland, 2009).

Researchers have also attempted to design team performance appraisals in the context of total quality management (TQM) systems—systems in which management is responsible for the implementation of effective teams and in which team performance



always supersedes individual performance. Janz and Harel (1993) attempted to resolve the apparent conflict between traditional human resource management (HRM), operationalized in their study as a system that values individual performance over team performance—and TQM principles by developing a three-dimensional model to evaluate performance data. From this analysis the authors believed that the conflict between these two schools of thought could be resolved by an appraisal that measures the team's process over the team's results dimensions; these process dimensions included decision making, administrative, and communication dimensions (Janz & Harel, 1993).

Lam & Schaubroek (1999) continued this line of research by examining team performance appraisals in the context of process versus results and group versus individual oriented approaches. They wanted to determine whether the type of performance dimensions and the unit of analysis would affect employees' satisfaction with the performance appraisal, their perceived accuracy of the performance appraisal, their expectations for improving their performance based on the results of the performance appraisal, and their actual performance. The authors provided a training session to Hong Kong, resident junior, front-line supervisors and had them participate in a team-based task. After the supervisors completed the team tasks, they were either provided with individual or team-based performance appraisals that were process or results oriented. The authors found that a process oriented appraisal, regardless of whether it was individual or team based, had the most positive impact on the employees' satisfaction with the performance appraisals, their perceptions of accuracy of the appraisal system, and their expectation for improving their performance. Employees who

received feedback based on the process oriented appraisal experienced greater improvement in their actual performance in comparison to employees who did not receive any feedback. In sum, these researchers demonstrate the importance of separating team process oriented measures from team results oriented measures. Although practitioners are likely to use team results oriented appraisals for administrative decisions, it is beneficial for them to use process oriented measures as supplements teams' performances within organizations; practitioners must implement a process oriented system that allows employees to receive feedback from this system.

Çiçek and colleagues (2005) proposed a more holistic team performance appraisal that captured both the results oriented and process oriented measures. After reviewing the previous work of Janz and Harel (1993), these authors also attempted to resolve the apparent conflict between HRM and TQM practices by creating a team performance measurement model that consists of four components, team structure, inputs, processes, and outputs. They posited that team structure can be assessed by determining how much a team possesses characteristics, such as clear objectives, communication and conflict management, commitment and involvement, culture, and administration. Team inputs can be defined as a team's technological, financial, human capital, and physical resources—along with other factors—that teams need to complete their team tasks. Team processes can be defined as helping to contribute to the team's outputs—results-oriented variables that should be related to the fulfillment of organization and or clients' needs. The authors implemented the system within the healthcare industry and found it to be a useful tool for assessing the performance of a neurological sciences team. Although they applied the

system to this industry, the authors believed that the team performance appraisal system should be based on the mission and goals of the organization and they believed that teams should collaboratively create the measures and ways to collect the completed appraisal measures; but, they believed that management needs to check its proposed appraisal process to determine whether it is appropriate within its specific organizational context (Çiçek, Köskal, and Özdemirel, 2005). This performance measurement model has the potential to be very useful to organizations due to the participation of employees in the creation of the appraisal system. This model would enable employees to have a voice in the system.

#### **Implementation of team performance appraisals in different environments.**

Researchers have attempted to assess team performance in dynamic and complex environments. For example, some researchers found data from the National Confidential Enquiry into Maternal Deaths that attributed poor communication and teamwork skills to the primary cause of subpar obstetric care (Morgan, Pittini, Regehr, Marrs, & Haley, 2007). Based on this observation, the reliability and validity of team performance appraisals was examined in a simulated obstetric environment (Morgan et al., 2007). Thirty four healthcare professionals (6 obstetricians, 16 nurses, and 6 anesthesiologists) from a single academic institution were assigned to one of four conditions that were intensive medical situations requiring hands-on activity within the obstetric teams. All teams completed the Human Factors Rating Scale (HFRS) team performance or the Global Rating Scale (GRS) of overall team performance. These healthcare professionals' performance was evaluated by nine external raters who were healthcare professionals.

The HFRS was an adaptation of the Operating Room Management Attitudes' Questionnaire (ORMAQ) that assessed team leadership structure, confidence-assertion, distribution of information within the team, team orientation, and error. The authors found that the aggregate intraclass correlation coefficient (ICC) for the external judges was reliable whereas the single rater ICC for external judges was low. Additionally, the authors did not find support for the validity of the HFRS measure whereas they found moderate support for the GRS. They suggested that the latter instrument not be used solely for the summary and feedback purposes rather than for formal evaluation purposes (Morgan et al., 2007).

Daradoumis, Martinez-Mones, and Xhafa (2004) examined team performance appraisals in a virtual teams context by conducting a case study on an "Application of Information Systems to Business" course that involved collaborative distance learning. Rather than assessing team learning outcomes with either quantitative measures or qualitative measures, these authors decide to use a hybrid of social network analysis (SNA) and descriptive statistics to formally assess workgroups in a computer-supported collaborative learning environment (CSCL). SNA consists of a social network, a group composed of persons or groups, termed as "nodes," which are connected by various interdependencies, such as beliefs, values, interests, and common knowledge. SNA attempts to determine the relationships between these different nodes to capture any meaningful interdependencies that may exist between individuals or groups. The students formed virtual teams at the beginning of the course and they completed specific assignments that pertained to the judgment and decision making tasks. All of the group

activity occurred on a Basic Support for Cooperative Work (BSCW) system. Based on their analyses, the authors concluded that some members who were influential in the general workspace were influential within the team, which led to their teams performing better than team members who were not influential in the general workspace (Daradoumis et al., 2004). This study is useful because it provides educators with a hybrid of SNA and quantitative measures for tracking and evaluating team processes and outcomes for virtual teams.

Some researchers have attempted to examine how well team performance appraisals used in an academic environment correlate with appraisals of team performance in workplace settings (Keefe, Glancey, & Cloud, 2007). Keefe et al. (2007) wrote an empirical review on the lessons learned from the implementation of a team performance appraisal for engineering students enrolled in a senior mechanical engineering course responsible for designing an industry-sponsored project. On average, this course contained 12 to 14 student teams of three to four members. These authors designed their team performance appraisal based on three dimensions of team performance: synthesis of a valid concept, management of resources, and interpersonal interaction and communication. Instructors and industry sponsors assessed the student team projects using these same dimensions. Assessment data were collected during a five year period on student teams' industry sponsored projects, and the researchers found that faculty tended to focus more on assessing team processes whereas industry sponsors focused on the team's results. Surprisingly, they found that faculty's assessment of students' team projects were not indicative of the industry sponsor's evaluation—a

difference which they attributed to industry's focus on results oriented dimensions (Keefe et al., 2007). This study demonstrates the importance of separating team process dimensions from team results oriented dimensions when one assesses team performance within organizations.

**Variables that influence effectiveness of team performance appraisals.** Some researchers have attempted to design and implement team performance appraisals, other researchers have examined variables that can hinder the effectiveness of these team performance appraisals. Similar to results reported for individual-based performance appraisals, some researchers have found that personality variables can affect team performance appraisals; specifically, their accuracy. Some researchers have investigated the role that the Big Five personality dimensions, specifically conscientiousness and agreeableness, play in rating leniency in team performance appraisals (Bernardin, Cooke, & Villanova, 2000). Bernardin and colleagues (2000) hypothesized that individuals with high levels of agreeableness were more likely to provide lenient team performance appraisal ratings; individuals with high levels of conscientiousness would provide the most accurate ratings; and individuals with low levels of conscientiousness and high levels of agreeableness would provide the most lenient team performance appraisal ratings. Based on peer evaluations from 111 students working on group projects concerning HRM issues (e.g., age discrimination), the authors found support for their hypotheses.

Self-monitoring has also been a variable of interest for researchers examining influences on team performance appraisals. Self-monitors are very aware of their

environment, they are sensitive to verbal and nonverbal cues from others, and they attempt to adjust their behavior to the situational context (Miller, 2001). Miller and Cardy (2000) sought to clarify the relation between team performance appraisals and self-monitoring, and they used self, peer, and supervisor team performance appraisals. Based on a laboratory study of students in groups working on case analyses and based on survey data from project teams across various occupations, the authors found that high self-monitors' team performance appraisals exhibited leniency and lacked agreement with other rater sources, and that they exhibited less convergence across rater sources (Miller & Cardy, 2000). Miller (2001) conducted a follow up exploratory study in which she examined the relation between self-monitoring and team appraisal satisfaction. Based on data from the same employees of 12 project teams, the author found that self-monitoring was negatively correlated with team appraisal satisfaction. The author argued that HSM team members were not likely to use the feedback to improve their future performance (Miller, 2001). If the purpose of team appraisals is to provide feedback on a team member's performance in hopes that he or she can improve the performance, these results seem very problematic for the success of the team performance appraisal system.

Aside from personality variables, demographic variables have been examined in the context of attributional biases in team performance evaluations. Wallace and Hinsz (2009) wanted to see how participants would attribute their performance on individual and group tasks. Ninety six undergraduate students were assigned to four-person teams and they completed successive card-sorting tasks. Although team members completed the same card-sorting tasks independently, their scores were combined to form a group score.

Twenty four students completed individual card-sorting tasks and received scores on their performance accordingly. After each task, participants in the contrived feedback condition were given feedback about their performance in comparison to previous individual or group (hypothetical) performance. Participants in the no comparative feedback condition did not receive this information. After the last trial, participants completed measures about the extent to which they believed their performance (on individual or group tasks) was a result of their own ability or based on external factors, such as luck. The authors found that group members made inflated internal and external attributions about their performance on the card-sorting tasks in comparison to individuals' evaluations of their own performance (Wallace & Hinsz, 2009). Additionally, they found that group members made inflated internal and external attributions about another hypothetical group's performance compared to individuals' assessments of similar hypothetical individuals' performance on the same tasks. The implications from these findings may be doubly problematic in an organizational context. First, if members of workgroups receive poor ratings on a team performance appraisal, they may be likely to disregard the appraisal information and attribute their poor performance to external factors. Second, if these members receive satisfactory or excellent team performance ratings which could be used to make administrative decisions, they may be more likely to want greater rewards because of the fact that they believe that team's performance was based on their effort and ability. Hence, researchers and practitioners should be mindful of the role that attribution bias plays on team performance appraisals.



Researchers have looked at the role that the delivery of the team performance plays in team performance appraisals. Fox, Bizman, and Garti (2005) wanted to determine whether a distributional appraisal method (DAM) was more effective than a traditional appraisal method (TAM) within a team context in terms of interrater agreement, ratee and dimension differentiation, and leniency. In comparison to a traditional performance appraisal, a distributional performance appraisal assumes that employees' performance levels on particular dimensions fluctuate and this appraisal attempts to incorporate these fluctuations into employees' evaluations. Nine teams of five members from a computer software organization rated each other using the DAM and TAM based on their performance on organizational assignments. The authors found that DAM scores had higher interrater agreement, specifically stereotype accuracy and differentiation, than TAM scores. Although the DAM scores exhibited higher interrater agreement, both measures exhibited leniency; but, as they noted, the authors lacked true scores (e.g., supervisor ratings) in which to compare the project teams' appraisal ratings (Fox et al., 2005).

Researchers have considered group communication modality in relation to the accuracy of team performance appraisals. Weismand and Atwater (1999) examined the amount of bias that exists in self and peer team-based performance appraisals based on whether team members interact with each other online or face to face. Students were assigned to groups that completed decision tasks online ( $N = 64$ ) or face to face ( $N = 27$ ). The authors found that actual team member contributions explained a large portion of variance in the peer appraisals for members assigned to teams that met online, and that

interpersonal affect accounted for the variance in peer evaluations for participants that met with their team members face to face. In a similar study, Kurtzberg, Naquin, and Belkin (2005) wanted to see if any differences existed between electronic-mail (email) and pen-and-paper peer-based team performance appraisals. Student teams completed labor-management team negotiations and either completed the team appraisals via pen-and-paper or by an email attachment (online). Over three studies, the authors found that participants who completed the email attachments of the team appraisals tended to give lower ratings of their team members than participants who completed the appraisals via pen-and-paper. The findings of the Kurtzberg et al. (2005) study align with the Weismand and Atwater (1999) study because the authors attributed the negative ratings to a reduction in participants perceptions of social obligation—which could have caused these participants to give higher ratings on the team appraisals. It is important to note that although these studies suggest that electronic performance appraisals and interactions among group members can reduce biases in performance appraisal, some researchers have shown that electronic communication can reduce team members' perceptions of performance and reduce perceptions of effective team processes (Fletcher & Major, 2006).

### **Self and Peer Team-based Performance Appraisals**

**Self-other comparisons.** Some recent research evaluates the correlation between self-and other team-based performance appraisals. Karakowsky & McBey (2001) examined this self-other relation regarding the influence of imputed expertise, operationalized as one's level of task competence. Undergraduate business students were

split into 36 teams and assigned to develop a strategy for negotiation that pertained to two business cases. After completing the negotiation strategy, the participants completed a measure of their self-perceived value of team contributions. Six human resource professionals, who served as expert judges, watched videos of the teams designing their negotiation strategies, and assessed their performance. Based on the results of these data, the authors found that the participants' self-perceived value of team contributions was significantly higher than the external judges' assessment of team members' involvement on the tasks. The findings suggest that researchers should consider team members' perceived competencies on tasks and the role that these perceptions have on team members' ratings of their significant contributions to the team.

Bergee and Cecconi-Roberts (2002) examined the relation between self appraisals with instructor and peer evaluations of music performance in a team based context. Across two studies, 87 undergraduate music majors were assigned to groups of three students, dependent upon whether they played an instrument or provided a vocal performance. They had four practice sessions in which they performed solos—which were recorded on videotape—in front of their group members. After their performance, all of the group members reviewed the videotaped solo performances, collectively discussed the performances, and then evaluated their own and others' performances. These weekly sessions led to an eventual final solo performance for each participant in the presence of their instructors. The participants, along with the instructors, evaluated their final performances. It is important to note that the instructors evaluated the performance of group members during the four prior practice sessions. Consistent with

previous research on self-appraisals, the authors found that participants' self-appraisals were not significantly correlated with instructors' evaluations of their performances. In fact, they found that over time, self-performance appraisals increased over time. Additionally, the authors found that although peer and instructor evaluations positively correlated with each other after the first session, this correlation declined over time. Throughout the proceeding sessions, group members' evaluations of their peers' performance were higher than instructors' evaluations.

Using a Social Relations Model, which captures the interdependencies between raters' assessments, Greguras and colleagues (2007) examined self- and peer team performance ratings. The authors tested the model on 29 organizational teams from various occupational settings. Team members assessed themselves and their teammates using a system for the multiple level observation of groups (SYMLOG) which captures the team member interdependencies. The SYMLOG assesses team members on teamwork orientation, individualism, rule compliance, dedication, affiliation, and motivation. In line with previous research (Harris & Schaubroeck, 1988), the authors found that self-ratings of SYMLOG dimensions were weakly correlated with peer ratings ( $r = .27$ ; Greguras, Robie, Born, & Koenigs, 2007) as they tended to be more lenient than peer ratings. Kwan, John, Robbins, & Kuang (2008) incorporated the SRM to determine characteristics of team members who self-enhance in their team performance evaluations. In one of three studies, the authors had 126 MBA students participate in group-decision making tasks in which they assessed their performance as well as their teammates' performance. Based on analyses of the SRM—and consistent with research on self-

enhancement in individual performance appraisals—the authors found that self-enhancers scored high on overt and covert measures of narcissism and they were judged by clinicians as hypersensitive, defensive, and less resilient (Kwan et al., 2008).

**Peer-other comparisons.** Along with self-other comparisons, researchers have also examined the relationship between peer ratings and other ratings in team performance appraisals. In one of the earliest studies, Saavedra and Kwun (1993) examined peer evaluations in self-managing work groups. Across three studies, over 350 business undergraduate and graduate students enrolled in an organizational behavior course, completed business assignments and case analyses in teams. Additionally, each member gave a class presentation. After completing these assignments, participants rated their teammates on their performance. The authors found that outstanding team contributors (conceptualized as over fifty-percent of their team members reporting these members as outstanding contributors) assigned the most accurate ratings of their teammates in comparison to average and below average team contributors. The authors posited that self-enhancement bias influenced average and below average team contributors' peer ratings of their teammates' performance because they rated their own performance higher than their teammates' performance. Saavedra and Kwun (1993) also found that average and below average team contributors tended to perceive the team peer evaluation system as unfair in comparison to outstanding contributors.

Morahan-Martin (1996) continued research regarding the reliability, leniency, and acceptance of peer appraisal ratings. One hundred and thirty six graduate students were placed in 32 debate teams and were evaluated by their teammates and their instructor.

Equal weights were placed on these evaluation scores to assign grades to the team members. The peer evaluation instrument consisted of questions concerning the team's performance, individual team members' performance, and the fairness of the team appraisal instrument. The author found that although the peer team appraisal was reliable, the participants rated the appraisal as fair, and that they were comfortable with the appraisal, the peer ratings exhibited leniency in comparison to instructor ratings on nine of the fourteen questions concerning presentation and research for both individuals and teams. Drexler, Beehr, and Stetz (2001) continued this line of research by examining how well peer appraisal could differentiate between individual performance on team-based tasks. Two hundred and ninety upper-level undergraduate students enrolled in an organizational behavior course were randomly divided into 56 teams that worked on business-related group assignments. After the completion of each assignment, participants rated their peers on their contribution to the team. The authors of this study found that the majority of the 56 teams did not differentiate among their peers' team contributions, and that these team members exhibited higher levels of distributive and procedural justice and team appraisal satisfaction than did the few teams that differentiated among their peers' team contribution.

Field studies have also examined the relationship between peer team appraisals with other appraisals. Kline (2001) assessed the extent to which team members' ratings of team performance correlated with supervisors' ratings of team performance. Seventy five employees who worked on 13 teams from different organizations completed a measure of team performance that was based on the ability of the team to solve problems, share the

workload, meet its objectives, and exhibit positivity; supervisors completed the same measure. From correlational analyses of the data, the author found that there was a lack of internal consistency for team members' assessment of team performance. Additionally, she found that the peer ratings of team performance did not correlate with the supervisors' ratings of team performance. A study on peer evaluations of team performance was also assessed in a clinical setting (Levine, Kelly, Karakoc, & Haidet, 2007). Levine and colleagues (2007) investigated the extent to which peer evaluations in a clinical clerkship correlated with traditional clinical assessments (National Board of Medical Examiners (NBME) score, clinic score, group readiness assurance test (GRAT), and individual readiness assurance test (IRAT)). Based on 152 students who worked in a team-based learning environment in a psychiatry clerkship, the authors found that the peer evaluation scores modestly predicted students' NBME, quiz, and clinical scores (Levine et al., 2007). It is important to note that when given the choice, 75% of the students chose not to have the peer evaluations count toward their grade. So, it could be argued that the peer evaluation's modest prediction of traditional clinical assessments could have resulted from the use of the evaluation, for developmental purposes.

**Peer ratings and organizational outcomes.** Although the aforementioned research shows problems with the use of peer appraisals in team-based settings across academic and organizational fields, there is ample evidence concerning the positive relation between the use of peer appraisals in teams and, arguably, organizational outcomes. In a repeated measures time-series design, Druskat and Wolff (1999) examined the effects and timing of developmental peer appraisals in self-managing work

groups. Using 294 undergraduates in 44 self-managing workgroups and 217 MBA students in 36 self-managing work groups, the authors found that the use of peer appraisals in teams can have an instant beneficial effect on members' perceptions of open communication, task motivation, social loafing, group viability, cohesion, and satisfaction. Additionally, these researchers found that over time peer appraisals can enhance interpersonal relationships within self-managing work groups as well as the teams' focus on the completion of tasks (Druskat and Wolff, 1999). In a similar study, Erez, Lepine, & Elms (2002) examined the role of peer appraisals on the functioning and effectiveness of 38 self-managed undergraduate teams that completed business case studies. They found that peer evaluations resulted in teams' levels of workload sharing, voice, cooperation, and performance that were higher than teams that relied on external evaluations (conducted by graduate students). In an Israeli kibbutz-owned manufacturing facility, Bamberger and colleagues (2005) conducted a longitudinal study on the effects of peer assessment on individual performance and team member contributions. The researchers found that peer assessments were correlated with an increase in supervisors' ratings of their subordinates' performance over time.

Dominick, Reilly, and McGourty (1997) wanted to determine if peer appraisals could be used as a feedback intervention to improve team member behavior. They conducted their study on 75 undergraduate and graduate students enrolled in an organizational behavior course who worked in teams to complete group decision making exercises. After each exercise, participants assigned to the feedback condition received feedback on their performance based on their peer (and self) evaluations of team



performance; the dimensions of team performance included collaboration, communication, decision making, and self-management which served as dimensions of team performance. Participants assigned to the exposure condition completed the same ratings but they did not receive any feedback from the ratings, and participants in the control condition neither completed peer (and self) ratings nor received feedback from any team performance ratings. Objective ratings of the participants' performance were made by experienced assessors blind to the experimental condition. Assessors were given standard instructions and rated participants on the five previously mentioned dimensions. The findings revealed that mere exposure to the peer evaluations led to significant improvements in team performance (based on the assessors' objective performance ratings) regardless of whether they received the feedback regarding the ratings (Dominick et al., 1997). So, it is possible that familiarizing participants with peer appraisals may cause them to correct any perceived weaknesses in their own team member behaviors.

### **Purpose**

Researchers have investigated the role that purpose plays in employees' ratings on individual and team-based performance appraisals. The purpose of performance appraisals is usually classified into two categories: administrative or developmental. Examples of administrative purposes are using appraisals to make hiring, promotion, pay raise, and disciplinary action decisions. In a university setting, administrative purpose can include using performance appraisals to inform students' final grades (Curtis et al., 2005). When performance appraisals are used for developmental purposes, they often are used to help employees monitor and improve their performance. It is important to note

that earlier studies (e.g., Harris, Smith, and Champagne, 1995) examined the effects of purpose on leniency in performance appraisals by looking comparing administrative and research purposes. A research purpose was operationalized as performance appraisals that were used for selection, specifically for validation studies. Based on this definition, the current review focuses primarily on studies that examine the use of performance appraisals for administrative and developmental purposes.

**Purpose and individual performance appraisals.** Much of the research on the effects of purpose on performance appraisals has dealt with individual performance appraisals. Jawahar and Williams (1997) decided to reexamine research conducted by Taylor and Wherry (1951) who hypothesized that performance appraisals would exhibit more leniency when they were used for administrative purposes compared to appraisals that were used for developmental purposes. Jawahar and Williams (1997) performed a meta-analysis on 22 previous studies that examined performance appraisal purpose, with a total sample size of over 57,000. The authors' results aligned with Taylor and Wherry's original findings because their results showed that performance appraisals ratings used for administrative purposes were significantly higher than ratings obtained for research or developmental purposes ( $d = .32$ ).

Shore and colleagues (1998) also examined performance appraisal purpose by investigating its affects along with self-appraisal information and feedback target on performance appraisal ratings. The authors hypothesized that participants' performance ratings would be more inflated when they were used for administrative purposes than when they were used for developmental purposes (Hypothesis 3). They also hypothesized

that when participants were told that they would have to discuss their ratings with the subordinates during feedback sessions, they were more likely to inflate their ratings if they were told that the appraisals would be used for administrative decisions than for developmental feedback (Hypothesis 4). When participants were told that they would have to explain their ratings of subordinates' performance to a high-level agent, the authors hypothesized that the participants were more likely to inflate their subordinate ratings if they were told the appraisals would be used for developmental purposes than when they would be used for administrative purposes (Hypothesis 4). 203 undergraduates acted as supervisors and were told to evaluate the performance of their subordinates' fictitious performance on a clerical task. These participants were told that the performance appraisals would be used as the basis for subordinates getting a research assistantship (administrative purpose) or for developmental feedback. The participants were also told that when they completed their appraisals, they would either have to provide one-on-one feedback to their subordinates or to the professor managing the study. The authors found that participants in the administrative purpose condition actually assigned lower ratings of subordinate performance than participants in the developmental feedback condition, but the difference was not significant (i.e., Hypothesis 3 was not supported). The results showed support for Hypothesis 4 because participants whose feedback target was their subordinate gave higher performance ratings when they were told these ratings would be used for administrative purposes than participants who were told the ratings would be used for developmental purposes. When participants' feedback target was the professor (i.e., high level organizational agent), their ratings were lower in

the administrative purpose condition than in the developmental feedback condition. These results suggest that when students are held accountable and have to provide justification for their ratings to a higher authority, they may be less likely to inflate ratings of their peers' performance when these ratings are used for administrative purposes.

**Purpose and team-based performance appraisals.** Although many studies have examined the effects of purpose on individual performance appraisals, the literature is scant on the effects of purpose on team-based performance appraisals. Instead, some authors have suggested purposes for which organizations should use team-based performance appraisals. Levy and Steelman (1997) provide a prototypical multi-rater source team appraisal system that includes self-, supervisor, and peer ratings. They believe that the purpose of the appraisal system serves as one contextual variable that should go into the overall team-performance appraisal process. For their prototypical team appraisal system, they suggest that team-appraisals be used for developmental feedback. They advise against using team-based appraisals for dual purposes, and do not believe that team appraisals should be used for administrative decisions because they argue that these decisions are often made at the individual level which is incongruent with the level of the performance evaluation (i.e., team). If organizations need to make administrative decisions, Levy and Steelman (1997) suggest that they design a team appraisal system that incorporates multisource ratings of job performance at the individual and team level along with multisource ratings on an administrative decision (e.g., promotion potential of a teammate).

London (2007) offers guidance to practitioners looking to assess group performance. In comparison to the previous authors, London (2007) suggests that consulting psychologists can use group or team-based performance appraisals for group development or evaluation. The author believes that when team appraisals are used for developmental purposes, practitioners can assess whether these teams are equipped with the right set of competencies and assess the overall cohesiveness of the team in terms of team members' cooperation with each other and their overall alignment to the groups' functions and goals. He also believes that team-based performance appraisals can be used for administrative purposes, including giving rewards to teams for their team performance or using the team appraisals for future sourcing and assignment decisions. Yet, similar to previous research on individual performance appraisals, the author suggests that quality and usefulness of the data may be influenced by the purpose of the ratings. He suggests that if team appraisals are used for administrative purposes, then employees may be more likely to inflate or be more lenient on their self-and peer ratings of team-performance. This indicates a need to test whether purpose can bring about leniency effects in team-performance appraisals.

### **Accountability**

Tetlock (1983) was one of the first researchers to examine the effects of accountability in social psychology. In this initial study, accountability was defined as the expectation of people to provide justification to someone else about their views, and the current definition of accountability is based on results from this study (Tetlock, 1983). In their review of research on the effects of accountability, Lerner and Tetlock (1999)

asserted that there are four specific dimensions of accountability: mere presence, identifiability, evaluation, and reason giving. Mere presence refers to the expectation that a person will observe one's behavior regarding their performance; identifiability refers to participants belief that they will be required to provide proof that they conducted ratings; evaluation refers to participants' belief that their behavior will be assessed by somebody else; and reason giving refers to the participants expectation that they will be called upon to provide an explanation for their actions (Lerner & Tetlock, 1999).

According to Lerner and Tetlock (1999), eight different types of accountability exist: accountability to an audience with known views, accountability to an audience with unknown views, predecisional accountability, postdecisional accountability, outcome accountability, process accountability, legitimate accountability, and illegitimate accountability. Accountability to an audience with unknown views refers to the fact that participants will be expected to provide justification for their performance behavior to an audience whose views are unknown as compared to the contrary with accountability to an audience with known views. Predecisional accountability refers to participants being told that they will be called upon to provide justification for their decisions prior to making them whereas postdecisional accountability refers to participants being told after they have made a decision that they justify their rationale for their decisions (Lerner & Tetlock, 1999). Outcome accountability refers to the effectiveness of participants' decisions being the primary criterion for their performance evaluation whereas process accountability refers to participants' decision processes being the criteria for their performance evaluation (Simonson & Staw, 1992). Legitimate accountability refers to

participants' belief that they feel obliged to provide justification to a source because they feel that that source should be obeyed, whereas illegitimate accountability entails participants holding a view contrary to legitimate accountability (Tyler, 1997). It is also important to note that Tetlock and Kim (1987) manipulated accountability in terms of preexposure-accountability and post-exposure accountability; preexposure-accountability refers to participants justifying their initial impressions of test-takers prior to receiving the test-takers' responses versus postexposure-accountability referring to participants justifying their impressions of the test-takers after they have received the test-takers' responses (Tetlock & Kim, 1987). Also, Harris (1994) conceptualized accountability in terms of upward versus downward accountability. Upward accountability refers to employees providing justification of their ratings of a subordinate to their supervisor, whereas downward accountability refers to employees providing the actual subordinates with justification for their ratings of those subordinates (Harris, 1994). Given these many different operationalizations of accountability, it is important to examine research on the effectiveness of accountability on reducing biases.

As mentioned earlier, Tetlock was one of the first social psychologist to study accountability and in one particular study (1983), he examined the effects of accountability on people's stances regarding social issues. It was hypothesized that participants in the accountability conditions would engage in more cognitive thought processing; it was also hypothesized that participants in the accountability condition who were expected to report to an individual with known social views would be more likely to shift their own views to align with those individuals and engage in less thought

processing; participants who were expected to justify their views to individuals whose views were unknown were more likely to engage in more cognitive thought processing so that they would be able to justify their decisions to individuals regardless of the stance that those individuals took. Forty-eight participants described their opinions on three social issues—affirmative action, capital punishment, and temporary issues—and the participants were assigned to one of four conditions: accountability to an individual with liberal views, accountability to an individual with conservative views, accountability to an individual with unknown views, and anonymity of their thought processes. The results indicated that participants who reported to individuals with known views were more likely to shift their stances on the social issues to align with the views of that individual; but when participants were accountable to individuals with unknown views, participants were more likely to engage in more cognitive thought processing (Tetlock, 1983). From Tetlock's (1983) initial study, it seemed as if accountability was most effective when people had to justify their views to people with unknown views.

Tetlock and Kim (1987) extended the previous line of research by examining the effects of accountability on participants' cognitive processing on a personality task. The researchers hypothesized that participants' levels of confidence would decrease on items on a personality measure that could be argued as either having true or false predictions. Sixty undergraduate students were told that they would participate in a person-perception process—how people created opinions of others based off of certain types of information. They were given Personality Research Form (PRF) responses from three persons for whom they were to write a biographical description based on their PRF responses. After



they completed this task, participants predicted the three test-takers' likely responses to additional sets of PRF questions and rate their level of confidence of whether their predictions would be true (likely to occur) or false (unlikely to occur). Participants were assigned to one of three accountability conditions: preexposure-accountability (participants described how they formed impressions of the test-takers' prior to receiving their PRF scores), postexposure accountability (participants explained how they formed impressions of the test-takers and how they wrote their test-takers' biographies after receiving the test-takers' PRF scores), and no accountability condition; participants in the accountability condition were told that their interviews would be audiotaped for data-analysis purposes. The results supported the original hypothesis, that participants' levels of confidence would decrease for PRF items that could either have true or false predictions and participants engaged in more cognitive thought processing. This research supported existing literature that suggested that accountability may be most effective when participants are required to justify their behaviors to an individual with unknown views but this occurred more in the preexposure accountability condition than in either the postexposure- or no accountability conditions.

Antonioni (1994) took a different approach to the study, applying the previous research to the workforce. Antonioni (1994) studied the effects of feedback accountability on upward appraisal ratings because the author was interested in whether employees assigned different ratings to their managers based on whether they were held accountable. He focused on a specific dimension when defining accountability, identifiability. Identifiability entailed subordinates writing their names on the upward

appraisals. It was hypothesized that managers who knew the names of subordinates rating them would view the appraisal process more positively than managers who received appraisals from anonymous subordinates. It was hypothesized that subordinates who identify themselves on their appraisals of their managers will have a less positive view about the upward appraisal system than subordinates who anonymously appraised their managers; further, it was hypothesized that subordinates who were accountable for their upward appraisals of their managers would make more positive ratings than would subordinates who anonymously provided upward appraisals of their managers. Thirty-eight managers and 183 subordinates participated in the study, and these participants were either assigned to the accountability condition or the anonymity condition. Subordinates completed the Upward Leadership Behavior Assessment (ULBA) of their managers in either condition, and managers received either a summary of the ULBA report from the anonymous subordinates or they received completed ULBA reports from accountable subordinates. The results supported the original hypotheses, that managers would support the upward appraisal system more when subordinates were accountable for their ratings, that subordinates in the anonymity condition would feel more positive about the upward appraisal process when they were anonymous (than accountable), and that accountable subordinates were more likely to significantly inflate their ratings of their managers compared to anonymous subordinates. Although this study would seem to imply that accountability would cause more leniency amongst subordinates in the appraisal systems within organizations, caution should be warranted for the author's

limited operationalizations of accountability as it entails only one of the four dimensions of accountability, identifiability (Antonioni, 1994).

Mero and Motowidlo (1995) broadened the scope of accountability by investigating the effects of accountability on the accuracy and favorability of performance ratings. It was hypothesized that raters who are held accountable for their ratings in a motivational context in which there are no special pressures to achieve a certain rating outcome will rate more accurately than will raters in the same motivational context who are not held accountable. Further, the authors hypothesized that some motivational contexts that do exert special pressures on raters to achieve certain outcomes. Accountable raters in these situations will feel the personal implications of their ratings more acutely than will nonaccountable raters and should be more motivated to avoid personal consequences that might be aversive for them. In comparison to Antonioni's (1994) limited definition of accountability, Mero and Motowidlo (1995) operationalized accountability as the participants being informed that they would be required to justify their ratings to the researchers. Two hundred and forty seven undergraduate students performed an in-basket task and watched a videotaped simulation during two sessions spanning two-weeks. The videotape contained vignettes that showed information about 4 simulated subordinates' performance. After assigning the ratings to the simulated subordinates, participants in the accountability condition provided their ratings to their supervisors (the researchers) in either a motivational or nonmotivational context; or participants were assigned to the nonaccountability condition. The researchers assessed participants' accuracy by creating a variation of the ratio of positive and

negative performance vignettes of the subordinates' performance. Both hypotheses were supported in this study, as participants who were held accountable with no motivational contexts rated the simulated subordinates more accurately than did nonaccountable participants. Also, participants who were held accountable with a motivational context, specifically that the subordinates received low performance ratings in the past, were more favorable on their ratings as compared to raters who were not held accountable (Mero & Motowidlo, 1995). The authors showed the potential positive benefits of the use of accountability within the appraisal system, mainly that employees possibly would provide more accurate ratings.

Frink and Ferris (1998) continued to examine the effects of accountability by looking at the relation between goals and accountability on performance in a laboratory and field setting. The authors hypothesized that participants would set higher goals in a high-accountability condition as compared to a low or no accountability condition. They would set higher goals as a type of self-handicapping strategy to attribute their poor performance to external outcomes (H1). It was hypothesized that participants would have higher levels of task attentiveness (H2a) and context attentiveness (H2b) in the high accountability condition in which the task outcomes are the primary criteria of accountability as compared to participants in the low or no accountability condition. Further, it was hypothesized that performance would be influenced by the interaction between accountability and goals where the correlation between goals and performance in the high accountability condition would be significantly different than in the low (or no) accountability condition. The authors hypothesized that this correlation would be less

salient in the high accountability condition than in the low (or no) accountability condition (H3). In the laboratory experiment, 115 undergraduate students were assigned to either a high or no accountability condition. Participants first completed a felt accountability questionnaire followed by a practice set of anagrams and math problems. They then completed a second questionnaire that instructed them to set goals for their performance on the final trial. Following the second questionnaire, participants completed the final trial of problems and completed a third questionnaire where they self-assessed their context and task attentiveness (during the final trial). While participants completed this final questionnaire, they were told that they could either leave after completing the questionnaire (no accountability) or that they would meet with a team leader to discuss their goals (accountability). All hypotheses were supported except for H2a and H2b. Participants in the accountability condition reported lower levels of task and context attentiveness than participants in the no accountability condition. Along with a laboratory experiment, the authors conducted a field study in which 27 telemarketers completed the same questionnaires as the participants did in the laboratory experiment. Rather than completing math and anagram tasks, the telemarketers performed their regular job duties and used calls per hour and revenue per hour to set their goals. To assign telemarketers to a low or high accountability condition, the authors performed a median split on the telemarketers' responses to a felt accountability questionnaire. Similar to the laboratory experiment, all of the hypotheses were supported except for H2a. Participants in the high accountability did not exhibit significantly different levels of task attentiveness than participants in the low (or no) accountability condition. One

implication from this study is that researchers and practitioners should be cautious in using accountability in goal-setting exercises with employees because it may elicit impression management and limit their ability to improve their performance. Another possible implication from this study is that if employees exerted higher levels of context attentiveness when performing their jobs after completing the felt accountability questionnaire, then employees who receive an accountability manipulation only when evaluating their and others' performance may use more cognitive processing when rating themselves and others on dimensions of job performance. This higher level of cognitive processing may make them less likely to assign unwarranted higher ratings of themselves and others on dimensions of job performance.

Beckner, Highhouse, and Hazer (1998) conducted a field study that examined the effects of upward accountability and rating purpose on peer-rater inflation and delay. They defined upward accountability as raters' expectation that they would provide justification for peers' ratings to their supervisor. Ninety three clerical, technical, client service, and administrative employees completed a peer-appraisal instrument and they were randomly assigned to a 2 (upward accountability versus no accountability) by 2 (administrative purpose versus research purpose) experimental design. The results showed that when workers were held accountable to their supervisors and conducted the peer-appraisals for research purposes only, they were more likely to delay their ratings. The authors found no significant effect for administrative purpose on rater delay, and they also found that purpose had no significant effect on peer-rating inflation. Surprisingly, the authors found no significant effects of upward accountability on peer-

rating inflation (Beckner et al., 1998). Although they did not find a significant effect for upward accountability on peer-rating inflation, upward accountability may be more effective for subordinates and managers providing ratings of each other.

Curtis, Harvey, and Ravden (2005) examined the effects of contextual variables that influence performance appraisal systems: appraisal purpose and rater accountability. The authors believe that participants who completed performance appraisals for administrative purposes would have more lenient ratings than participants who completed appraisals for developmental purposes (Hypothesis 1). They also hypothesized that participants who were told that they had to justify their ratings to the experimenter (upward accountability) would have less inflated performance ratings of a confederate than participants who were not held accountability; it was also hypothesized that participants who had to justify their ratings to the confederate (downward accountability) would have more lenient ratings than participants who were not held accountability. Therefore participants in the upward accountability condition would have less lenient ratings than participants in the downward accountability condition (Hypothesis 2). The third hypothesis stated that there would be an interaction between purpose and accountability such that participants in the downward accountability condition would provide the most lenient performance ratings when they were told they would be used for administrative purposes; whereas participants in the upward accountability condition would provide the least lenient ratings when they were told that their ratings would be used for developmental feedback. 133 undergraduates assessed a confederate on a fictitious sales call audiotape and this audiotape was of moderate to low quality (based on

the same participants' average rating of the audiotape without any experimental manipulations). Overall, the authors found that when participants were told that they would be held accountable to the experimenter and that their ratings would be used for developmental purposes, they provided less lenient ratings across experimental conditions. In addition, when participants were told that they would be held accountable to the confederate and that their ratings could be used to make administrative decisions, they provided the most lenient ratings in comparison to the other experimental conditions (Curtis et al, 2005). So these results show support for the role that upward accountability can play in reducing leniency in performance appraisals.

Gordon and Stuecher (2001) extended the former line of research on upward accountability by examining the effects of accountability and anonymity on the linguistic complexity of teaching evaluations. The authors hypothesized that participants who described their instructor evaluations to a faculty member (i.e., upward accountability) would have more complex evaluations, and participants who described their instructor evaluations to another student, those evaluations would exhibit less complexity. The authors also hypothesized that the condition in which participants provided an explanation of their evaluation to a faculty member along with signing the form would yield the most complex instructor evaluation. One hundred and fifty seven undergraduate students were assigned to a two (anonymous, signed) by 3 (low student, high student, and high-faculty accountability) between subjects design. The authors found that there was no significant main effect for anonymity on the complexity of teaching evaluations but they found that participants in the high-faculty accountability condition exhibited the



highest complexity on their evaluations of their teachers as compared to low and high student accountability conditions. As predicted, the authors also found that the condition in which upward accountability was coupled with an identifiable form yielded the most evaluation complexity as compared to the other conditions. Several factors from this study should be noted, that the authors were using anonymity and accountability as two independent variables. This should be of interest to the reader because accountability entails a form of anonymity within the definition, mainly identifiability. Also, in contrast to the Beckner et al. (1998) study and in conjunction with previous accountability research, this study's upward accountability condition did help aid in predicting more complex instructor evaluations, which could be argued to produce more accurate student evaluations of instructors.

Instead of looking at the effects of accountability based on previous conceptualizations, Brtek and Motowidlo (2002) examined the effects of accountability in terms of procedure and outcome accountability on interview validity. Three hundred and thirty eight undergraduates were assigned the role of an interviewer and they were assigned to one of four conditions: procedure accountability only, procedure and outcome accountability, outcome accountability only, and no accountability. The authors hypothesized that holding participants procedurally accountable would increase the validity of the interview, whereas holding participants accountable for accuracy of the outcome of their interview ratings would lower the validity of the interview. It is also important to note that the positive effects that procedure accountability had on interview validity was mediated by participants' attentiveness (Brtek & Motowidlo, 2002).

Roch (2005) examined the effects of rating audience and financial incentives—two motivational factors that some researchers believe relate to accountability—on performance ratings. The author decided to study financial incentives based on performance appraisal research showing that consequences are needed to make employees feel accountable for their ratings and for following a set of rating procedures. One hundred and forty nine undergraduate students rated videotaped performances of persons in an assessment center group exercise and were assigned to conditions based on level of financial incentive and audience characteristics (expert, ratee, dual). Based on these performance ratings the author found that audience characteristics, a core component of accountability, could curb rater leniency in the face of financial incentives because participants in the dual and expert audience conditions had lower ratings in comparison with participants in the ratee audience condition (Roch, 2005).

In a follow up study, Roch and McNall (2007) investigated the influence of accountability on performance ratings and whether other factors may impact accountability and performance ratings. The independent variables in this study were audience type (no audience, peer, instructor; peer and instructor) and identifiability (writing name on instructor rating sheet, not writing name on instructor rating sheet). Audience type was operationalized as participants reading written instructions stating they would have to present their ratings to one of the four previously mentioned audiences. 315 undergraduate students completed two questionnaires. On the first questionnaire, students assessed their “effort/importance” and pressure (e.g., “I feel very tense as a student”). The second questionnaire contained the different conditions,

instructor evaluation items, demographics, and a measure of felt accountability (“I felt accountable for my ratings”). As part of the second questionnaire, students completed one of eight different instructor evaluation forms based on the different experimental conditions (Audience x Identifiability). Based on the results of a hypothesized structural model, students with higher levels of felt accountability assigned higher ratings to their instructors’ performance; and the more participants believed in the importance of being a student, the more they assigned higher ratings to their instructors. If students had to write their names on the instructor rating forms, experienced pressure while rating the instructors, and felt general pressure being a student, they were more likely to assign lower ratings to their instructors (Roch & McNall, 2007). Current researchers that use accountability do not operationalize it as felt accountability rather than including dimensions such as identifiability. So, this research shows the potential benefits of using accountability in performance appraisals systems, mainly to curb leniency biases.

Mero, Guidice, and Brownless (2007) studied the effects of audience and form of accountability on rater response and rater behavior. One hundred and ninety seven MBA students performed group tasks, and provided ratings to either the team of students (downward accountability) or to the session administrator (upward accountability); it is important to note that they also had a mixed accountability condition in which the participants justified their ratings to the session administrator and to the team of students (Mero et al., 2007). The participants also justified their ratings to each audience either by face-to-face or by justifying their ratings via a written evaluation of why they believed the team of students received the ratings that the participants assigned to them. The

authors found that when the participants provided their ratings of the student team's performance to the session administrator, these ratings were less inflated and less lenient in comparison to the team of students or a no accountability condition; the authors also found that raters who were required to meet face-to-face to provide justification for their ratings were more accurate than participants who had to provide a written justification of their ratings (Mero et al., 2007). This study aligns with existing research regarding how accountability can reduce inflation and leniency in performance appraisals to yield more accurate ratings.

### **Summary: Literature Review**

**Team-based performance appraisals: Self- and peer ratings.** Based on the current review, there is a need to continue to look for ways to help incorporate self- and peer ratings more effectively into current performance appraisals systems. Specifically, there is a need to help raters reduce their tendency to overrate themselves and their teammates on dimensions of job performance. Many organizations are creating and revising job positions so that employees work in cross-functional and virtual teams to carry out their mission critical functions. As a result, many supervisors are unable to see their subordinates' job performance on a day-to-day basis. So, as this trend occurs, it is critical that organizations have team-based performance appraisal systems in place that can help determine whether their employees are working effectively across teams to execute job duties and responsibilities. These appraisals should contain peer ratings to determine how their subordinates are working together as a team to accomplish tasks. There is also a need to incorporate self-appraisals into these appraisal systems to let

employees have direct input into the appraisal process and increase perceptions of fairness from appraisal outcomes and satisfaction with the overall system and with the organization. When practitioners design these team-based performance appraisals, it is important to incorporate research on effective teams, specifically dimensions of effective team performance such as mutual performance monitoring and shared mental models. The current literature on teams-based performance appraisals is growing and researchers should look for ways to help employees better evaluate themselves and their peers on team-based tasks.

**Purpose.** Decades of research shows that the purpose of individual performance appraisals plays an influential role in students' and employees' performance ratings of themselves and their peers. When performance appraisals are used for administrative purposes (e.g., promotions, pay raises, final grades), people are more likely to inflate their self- and peer ratings compared to when performance ratings are used for developmental purposes (i.e., used as feedback to improve job performance). Regarding teams, some authors believe that team-based performance appraisals should not be used for administrative purposes. They believe that it would be inappropriate to use team-based performance appraisals for administrative purposes when these purposes are managed at the individual level (e.g., promotion of individual employees). Instead these authors believe that team performance appraisals should be used for developmental purposes, including team members assessing their own and teammates' ability to accomplish given tasks. Recent studies have examined the effects of accountability on reducing individuals' tendency to inflate their performance ratings when their appraisals

were used for administrative purposes and their results show the potential of accountability to reduce leniency effects in performance appraisals that are used for administrative purposes (Curtis et al., 2005).

**Accountability.** Although accountability is a buzzword used across many academic and occupational fields, this variable has been purported to improve performance appraisal systems by potentially reducing the leniency bias that still plagues many self and peer evaluations in individual and team-based performance appraisals. Despite the many different types of accountability, a definition should include most of Lerner and Tetlock's (1999) four dimensions (mere presence, identifiability, evaluation, and reason giving) of accountability. When accountability is defined this way, it may make employees provide ratings that better reflect their own and their peers' job performance.

### **Purpose: Current Study**

Due to the recent empirical investigation in which accountability was found to reduce leniency effects in self-appraisals on a job-related task (Smith & Switzer, 2009), the author of the current study wanted to determine if any literature existed that examined the effects of accountability on leniency reduction in team-based performance appraisals. To the author's knowledge, no studies have directly examined the effects of accountability on reducing inflation in peer and self team-based performance appraisals.

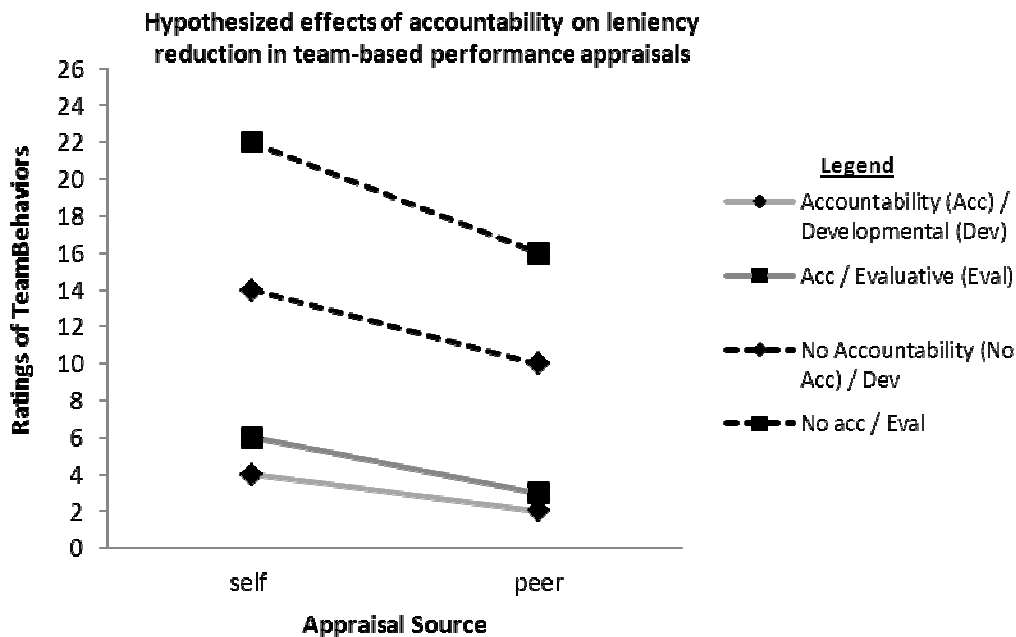
The present study attempts to extend the findings from the previous study of the effects of accountability on leniency reduction in individual-based performance appraisals to a team performance appraisal context. Specifically, the current study adds to

the existing body of research on performance appraisals by investigating whether accountability can reduce leniency bias on team-based performance appraisals, specifically for self- and peer ratings of team behaviors. The current study also adds to the existing body of literature on performance appraisals by determining whether the purpose (evaluative or developmental) can influence self- and peer ratings on team-performance appraisals. It is important to note that the evaluative purpose condition was the equivalent of an administrative purpose because it was operationalized as participants being told that they would have to complete the Team Behaviors Form to receive course credit for participating in the lab study, or as participants reading (in the field study) that their ratings on the Team Behaviors Form would be used as a basis for their own and their teammates' final grades. The present study applied accountability to a realistic setting because it dealt with team-based performance appraisals on projects that are reflective of projects in industrial and organizational settings, such as process control plant operations tasks and engineering projects. For the lab study, accountability was operationalized as the experimenter instructing the participants to provide written justification on the Team Behaviors Form (TBF) form for their ratings and telling them they would have to provide verbal justification for their ratings to a local nuclear process control plant manager during a future one-on-one session (upward accountability). The experimenter did not tell participants in the no accountability condition to provide written justification for their ratings and did not tell them they would have to verbally justify their ratings to a superior. For the field study, accountability was operationalized as participants reading instructions on the TBF that required them to provide written

justification on the form for their ratings and that indicated they would have to provide verbal justification for their ratings to a superior during a future one-on-one session with that superior (upward accountability). Participants in the no accountability condition were not required to provide written or verbal justification for their ratings. Throughout this paper, upward accountability is referred to as accountability. It is important to note that the current operationalization of accountability was strengthened because not only did participants read that they would have to report their ratings to a member of the higher audience, but also report to one of the higher-ups who has credibility (i.e., local nuclear process control supervisor and professor). This operationalization optimizes the saliency of accountability.

## Hypotheses

Figure 1. Hypothesized effects of accountability, purpose, and appraisal source on ratings of team behaviors.





H1: Participants in the no accountability condition would have significantly higher self-ratings of team behaviors than peer ratings when the appraisals were used for evaluative purposes.

H2: Participants in the no accountability condition would have significantly higher self-ratings of team behaviors than peer ratings when the appraisals were used for developmental purposes.

H3: Participants in the accountability condition would not have higher self-ratings of team behaviors than peer ratings when the appraisals were used for evaluative purposes.

H4: Participants in the accountability condition would not have higher self-ratings of team behaviors than peer ratings when the appraisals were used for developmental purposes.

Note that hypotheses H1-H4 taken together hypothesize a 2-way interaction such that there would be an effect for source (self-ratings > peer ratings) but only in the no accountability condition (and this is regardless of purpose).

H5: Participants in the accountability condition would have lower ratings of team behaviors than participants in the no accountability condition regardless of the purpose of the evaluation (i.e., a main effect for accountability).

H6: Participants would give higher ratings of team behaviors if the appraisals were used for evaluative purposes than if they were used for developmental purposes (i.e, a main effect for purpose).

## CHAPTER TWO

### METHOD

#### **Overview of Method Section**

The following section provides a detailed description of the method that was used in the development of the current experiment. The method section includes a description of the participants, apparatus, measures, pilot study, and a description of the procedures that were used in the current experiment. The experiment consisted of two studies, a field study and a laboratory study. The author felt that including a lab study in his experiment would serve as a clean way to test accountability because he could manipulate it in a controlled setting. The field study consisted of undergraduate students enrolled in different courses across a variety of majors including engineering, management, and computer sciences. The laboratory study consisted of undergraduate students who participated in a face-to-face two-person process control team.

#### **Method**

**Participants.** One hundred and twenty undergraduate students participated in the lab study and were enrolled in introductory Psychology courses at a midsize southeastern university. One hundred and eighty-one undergraduate and graduate students participated in the field study and were enrolled in courses across a variety of majors including management, engineering, computer sciences, and applied economics and statistics at midsize southeastern universities. Participants received course credit for participating in the study.

**Measures.** *Team Behaviors Form.* Participants assessed their own and others' performance on eight dimensions of team behaviors (team processes): mutual performance monitoring, backup behavior, adaptability, team orientation, shared mental models, closed loop communication, mutual trust, and team leadership. The participants rated themselves and each of their peers on team behaviors using a 5-point Likert scale anchored by 5 = Excellent, 1 = Poor (e.g., "How would you rate your awareness of individuals' progress?" or "How would you rate M1's awareness of individuals' progress?"). For mutual performance monitoring, participants were asked to rate, "Your awareness of individuals' progress." For backup behavior, participants were asked to rate, "Your willingness to take on the roles and responsibilities of individuals who need help." For adaptability, participants were asked to rate, "Your adaptation to challenges and unforeseen situations." For team orientation, participants were asked to rate, "Your willingness to work as a team." For shared mental models, participants were asked to rate, "The extent to which you are on the same page with other team members." For closed loop communication, participants were asked to rate, "Your communication with other team members." For mutual trust, participants were asked to rate, "Level of mutual trust that you have with other team members." For team leadership, participants were asked to respond "yes" or "no" to: "Do you feel that leadership exists within your team?". If they answer yes, they were asked to, "rate the effectiveness of your leadership abilities in the team."

The questionnaire was developed as follows. The initial questions were drafted based on the extensive literature review on team performance appraisals and based on the

importance of separating team process dimensions from team outcome dimensions. The dimensions of the form were based on Salas and colleagues' (2005) important dimensions of team effectiveness. After the questions were created, pilot studies were conducted in which students enrolled in an Education course completed the Team Behaviors Form. After students completed the form, interviews were conducted with the instructor of the course regarding the clarity, wording, and general structure of the Team Behaviors Form (and these interviews were taken into consideration if changes needed to be made on the form). Data from these pilot studies were collected and basic descriptive statistics were run to ensure that there were no irregularities in the self-rating questionnaire.

As part of the pilot study, we compared the Team Behaviors Form with the Team Member Assessment (TMA) form. On the TMA form, participants assessed their own and others' performance on five dimensions of team member effectiveness: contributing to the team's work, interacting with teammates, keeping the team on track, expecting quality, and possessing the relevant knowledge, skills, and abilities. The participants rated themselves and each of their teammates on a 5-point Likert scale anchored by 5 = Exceeds Expectations and 1 = Below Expectations (e.g., How well did this team member contribute to the team's work?). The TMA form was based on Loughry, Ohland, and Moore's (2007) short version of the Comprehensive Assessment of Team Member Effectiveness (CATME). The short version of the CATME contains 33 items that map on to five dimensions: contributing to the team's work (8 items,  $\alpha = .96$ ), interacting with teammates (10 items,  $\alpha = .95$ ), keeping the team on track (7 items,  $\alpha = .93$ ), expecting

quality (4 items,  $\alpha = .9$ ), and possessing the relevant knowledge, skills, and abilities (4 items,  $\alpha = .91$ ).

### **Pilot Study**

A pilot study was conducted in an Education course in which students assigned to three to five member teams completed the Team Behaviors and TMA forms. During the middle of the semester, participants were administered the Team Behaviors Form (TBF) and they were explicitly told that the form would be used for developmental purposes only. The instructions on the TBF were as follows, “This is a feedback form intended to improve your team’s effectiveness. You are being asked only for your team number so that your entire group can get feedback, but your name will not be associated with this form and anything you say will be anonymous. This form will NOT be used to evaluate or grade you or your team members. Please provide ratings for each item below as they pertain to the effectiveness of your team. Consider only project related behaviors.” After the participants completed the TBF, the professor collected the forms, sealed them in an envelope, and sent them to the experimenter. The experimenter, along with his colleagues, conducted basic analyses on these data looking for outliers, as well as examining the average scores among the items. Based on these findings, the TBF demonstrated a high level of reliability, but the author noticed that the ratings seemed to exhibit a leniency bias. Based on these pilot data, the author thought that accountability should be incorporated within the framework of this TBF. Therefore, based on this pilot study, two versions of the TBF were created: an accountability version and a no accountability version. The instructions for the accountability version of the Team

Behaviors Form were as follows, “Please fill in a rating for each of your team members beside their code (e.g., M1). After you complete the ratings, please provide written justification on the back of this sheet for your ratings, as you will be required to verbally justify your ratings to me during a scheduled one-on-one session next week. Please consider project-related behaviors only. Please fill-in each of your team members’ names below, including yourself; however, remember that your responses will remain anonymous.” The instructions for the no accountability version of the TBF were as follows, “Please fill in a rating for each of your team members beside their code (e.g., M1). Please consider project-related behaviors only. Please fill-in each of your team members’ names, including yourself; however, remember that your responses will remain anonymous.”

In addition to the Team Behaviors Form, students were administered the TMA form in class, at the end of the semester. During this class session, participants received feedback concerning their team’s progress, and they were allowed to discuss issues as these pertained to their team’s progress. After this discussion, the instructor handed students the TMA form and instructed them to not discuss their ratings with their other team members. There were two versions of the TMA form, an accountability version and a no accountability version. The instructor separated the teams based on the accountability condition. Before the students completed the TMA measure, the instructor told them to turn their form face down when they completed it (and then she would collect them). Students who received the no accountability version of the form received the following written instructions, “Please fill in a rating for each of your team members

beside their code (e.g., M1). Please consider project-related behaviors only. Please fill-in each of your team members' names, including yourself, however remember that your responses will remain anonymous." When they completed this form and turned it face down, they read the following instructions, "In the spaces below, elaborate, by person, on ALL the ratings you gave on the front of this page. Do NOT change any ratings you gave on the front." If the students were in the accountability condition, they read the following instructions, "Please fill in a rating for each of your team members beside their code (e.g., M1). After you complete the ratings, please provide written justification on the back of this sheet for your ratings, as you will be required to verbally justify your ratings to me. Please consider project-related behaviors only. Please fill-in each of your team members' names below, including yourself, however remember that your responses will remain anonymous." After they completed the form, they turned it face down and the professors collected the forms. The instructor collected all TMA forms and sealed them in an envelope. When the experimenter collected the forms, he conducted basic descriptive statistics and reliability estimates on these forms (as well as examining the data for any outliers). Additionally, the author examined difference scores for each item, based on whether students completed an accountability or no accountability version of the form.

In sum, based on the results of this pilot study, the experimenter concluded that an accountability manipulation needed to be incorporated within the TBF and that instructors needed to carefully review a script for the administration and collection of this form.



## **Design**

This study was a 2 x 2 x 2 factorial design in which accountability was a between-subjects 2-level independent variable (accountability and no accountability); the purpose of the appraisal instrument was a between-subjects 2-level independent variable (evaluative and developmental); and the source of the appraisal (rater source) was a between-subjects 2-level independent variable (peer and self). The dependent variable was a composite variable termed the “Big5,” which was the average of participants’ ratings across 5 dimensions on the Team Behaviors Form (TBF) that are considered as dimensions of effective team performance: mutual performance monitoring, backup behaviors, adaptability, team orientation, and leadership (Salas et al., 2005). Thus there was a “Big5” variable for self-ratings and a “Big5” variable for peer ratings.

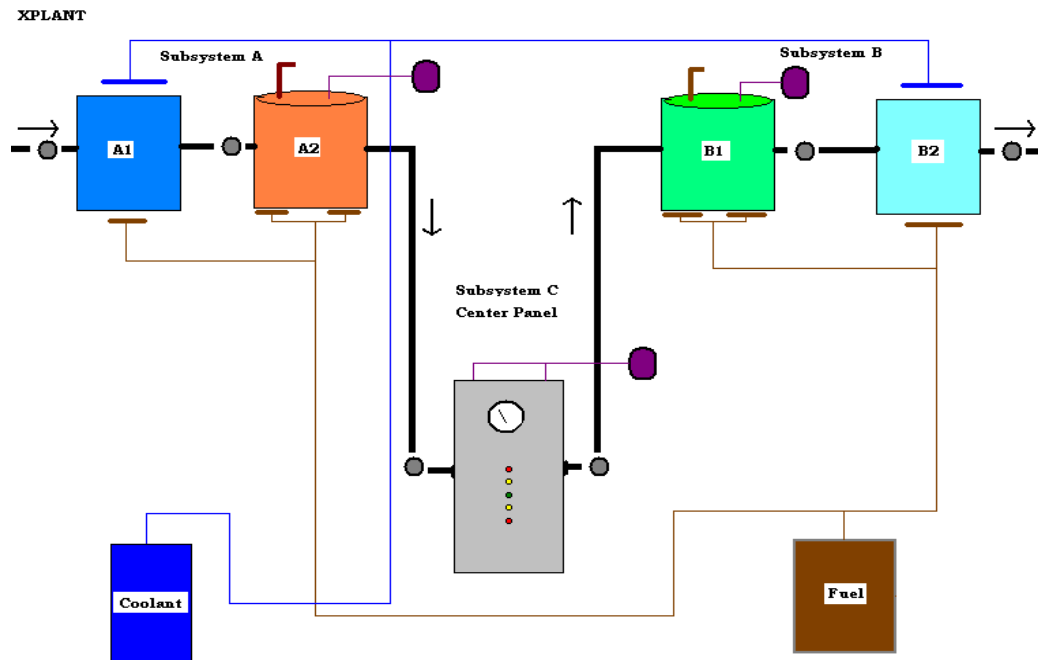
### **Study 1: Laboratory Study**

#### **Apparatus**

**Process Control Plant Simulator.** The laboratory study used a process control plant simulator that is based on simulator development by Switzer and Idaszak (1989) and Fjelde and Switzer (1994). This simulator contained three sections, subsystems A and B, and the center panel. Subsystems A and B were the primary subsystems that were shown on a computer screen, and participants had to control these subsystems by two keyboards that were placed in front of each screen. Each participant was randomly assigned to either subsystem A or B and each participant was seated in front of the corresponding computer screen and keyboard. The center panel contains a manual switch and LED display screen that is located between subsystems A and B. Although all

sections of the plant are connected, subsystems A and B control separate sections of this plant. The operator's primary task was to monitor the fluid levels in the 18 tanks in subsystems A and B; the fluid levels and tank pressures in the 3 center panel tanks; and to fix the parameters if they exceed specified upper and lower limits. Additionally, the operator's secondary task was to manage the input and output valves to these tanks to optimize fluid throughput. The functioning of the plant was based on a "fluid" processing system created so that operator A's main job responsibility consisted of managing the total input to the plant and the output of subsystem A to the center section (controlled by the center panel). Operator B controlled the flow from the center panel into subsystem B and was responsible for the total system output.

Figure 2. XPlant Process Control Plant Simulator Diagram



**Procedure.** Participants were recruited from the human participants for research (HPR) system. When participants used the HPR system to volunteer for the current study, they were given the length of time for the current study, between forty to forty-five minutes. When they entered the laboratory, they read and signed a consent form. After signing the consent form, the experimenter gave participants an overview of the study and told them that they would have to work in two-person process control teams. These teams received training on how to use the process control plant simulator and these team members participated in a practice session that lasted ten minutes. During the training, the experimenter told the participants that their primary task was be to monitor the fluid levels in the 18 tanks in subsystems A and B; the fluid levels and tank pressures in the 3 center panel tanks; and to fix the parameters if they exceed specified upper and lower

limits. Additionally, the experimenter told the participants that their secondary task was to manage the input and output valves to these tanks to optimize fluid throughput. The experimenter randomly assigned participants to be operator A or B on the two-person process control team.

The functioning of the plant was based on a “fluid” processing system created so that operator A’s (on the two-person process control team) main job duty consisted of managing the total input to the plant and the output of subsystem A to the center section (controlled by the center panel). Operator B controlled the flow from the center panel into subsystem B and was responsible for the total system output. After the completion of the process control tasks, the experimenter read the instructions listed on the Team Behaviors Form (TBF) to the participants. After listening to the instructions, the participants completed TBF where they rated their own and their team member’s team behaviors pertaining to the process control plant simulator.

Both participants were assigned to one of four conditions based on accountability (accountability and no accountability) and appraisal purpose (evaluative and developmental): accountability-evaluative purpose, accountability-developmental purpose, no accountability-evaluative purpose, and no accountability-developmental purpose. After the participants completed the Team Behaviors form, the experimenter collected these forms for further analyses.

**Accountability and purpose conditions.** If participants were in the accountability-evaluative purpose condition, they read the following instructions: “This is

a feedback form intended to assess team member behaviors. You MUST complete this form to receive credit for participation in this study.

Please fill in a rating for yourself and your teammate. After you complete the ratings, provide written justification on the back of this sheet for your ratings as you will be required to verbally justify your ratings to a local nuclear process control manager during a scheduled one-on-one session next week. You MUST meet with this manager next week to receive credit for participation in this study. Please consider behaviors only related to this process control task. Please fill-in each of your and your teammate's names below, however remember that your responses will remain anonymous."

If participants were in the accountability-developmental purpose condition, they read the following instructions: "This is a feedback form intended to assess team member behaviors. This form is for developmental purposes only.

Please fill in a rating for yourself and your teammate. After you complete the ratings, provide written justification on the back of this sheet for your ratings as you will be required to verbally justify your ratings to a local nuclear process control manager during a scheduled one-on-one session next week. Please consider behaviors only related to this process control task. Please fill-in each of your and your teammate's names below, however remember that your responses will remain anonymous."

If participants were in the no accountability-evaluative purpose condition, they read the following instructions: “This is a feedback form intended to assess team member behaviors. You **MUST** complete this form to receive credit for participation in this study.

Please fill in a rating for yourself and your teammate. Please consider behaviors only related to this process control task. Please fill-in each of your and your teammate’s names below, however remember that your responses will remain anonymous.”

If participants were in the no accountability-developmental purpose condition, they read the following instructions: “This is a feedback form intended to assess team member behaviors. This form is for developmental purposes only.

Please fill in a rating for yourself and your teammate. Please consider behaviors only related to this process control task. Please fill-in each of your and your teammate’s names below, however remember that your responses will remain anonymous.”

## **Study 2: Field Study**

**Procedure.** At the beginning of the study, the researcher contacted professors across a variety of majors and requested their permission to distribute the TBF in their classrooms. Professors were selected if they assigned team-based projects to their students. If they consented, the researcher then emailed them with specific instructions and the following attachments: Team Behaviors Form (accountability/developmental or accountability/evaluative), Team Behaviors Form (no accountability/developmental or no

accountability/evaluative), Protocol (Evaluative or Developmental), TBF Dimension Defined document, and Field Study IRB Approval form (see Appendices).

**Team Behaviors Form: Developmental Purpose.** If the researcher assigned a professor's classroom to the developmental condition, he emailed the following message to the professor:

“Good Morning,

I hope all is well. I spoke with you earlier about using my Team Behaviors Forms (TBFs) in your classroom to assess students' performance on a team-based project. I have attached two versions of the TBF: Developmental-Accountability (Version A) and Developmental-No Accountability (Version B). The protocol (attached) explains each condition and provides instructions on how to distribute the forms to your students at the end of the assigned project.

Before administering the TBF, please review the “TBF Dimensions Defined” document (attached) with your students; this document defines the different dimensions of effective team performance. Also, for your records, I am attaching an IRB authorization document stating that my experiment complies with local and Federal regulations.

Please let me know if you have any questions and I really appreciate your assistance!

Thank you—Brett”

At the end of the semester, professors randomly administered the Team Behaviors Form (TBF) to their students so that some students received the accountability (developmental) or no accountability (developmental) version of the form. The professors were instructed to attempt to randomly distribute the forms so that half of their students received the accountability (developmental) version and the other half received the no accountability (developmental) version of the TBF.

If the students received the accountability (developmental) TBF, they read the following instructions:

“This form will NOT be used to evaluate or grade you or your team members, it is for developmental purposes only.

Please fill in a rating for each of your team members beside their code (e.g., M1). After you complete the ratings, please provide written justification for your ratings as you will be required to verbally justify your ratings to me during a scheduled one-on-one session next week. Please consider project-related behaviors only. Please fill-in each of your team members’ names below, including yourself, however remember that your responses will remain anonymous.”

If the students received the no accountability (developmental) TBF, they read the following instructions:

“This form will NOT be used to evaluate or grade you or your team members, it is for developmental purposes only.



Please fill in a rating for each of your team members beside their code (e.g., M1). Please consider project-related behaviors only. Please fill-in each of your team members' names below, including yourself, however remember that your responses will remain anonymous”

After the students completed the TBF, the professors collected them and returned them to the researcher by one of three ways: hand-delivery, mail (e.g., USPS), or email.

**Team Behaviors Form: Evaluative Purpose.** If the researcher assigned a professor's classroom to the evaluative condition, he emailed the following message to the professor:

“Good Morning,

I hope all is well. I spoke with you earlier about using my Team Behaviors Forms (TBFs) in your classroom to assess students' performance on a team-based project. I have attached two versions of the TBF: Evaluation-Accountability (Version A) and Evaluation-No Accountability (Version B). The protocol (attached) explains each condition and provides instructions on how to distribute the forms to your students at the end of the assigned project.

Before administering the TBF, please review the “TBF Dimensions Defined” document (attached) with your students; this document defines the different dimensions of effective team performance. Also, for your records, I am attaching an IRB authorization document stating that my experiment complies with local and Federal regulations.

Please let me know if you have any questions and I really appreciate your assistance!

Thank you—Brett”

At the end of the semester, professors randomly administered the Team Behaviors Form (TBF) to their students so that some students received the accountability (evaluative) or no accountability (evaluative) version of the form. The professors were instructed to attempt to randomly distribute the forms so that half of their students received the accountability (evaluative) version and the other half received the no accountability (evaluative) version of the TBF.

If the students received the accountability (evaluative) TBF, they read the following instructions:

“This form will be used to evaluate you and your team members. This form will be used as a basis for you and your team members’ final grades.

Please fill in a rating for each of your team members beside their code (e.g., M1). After you complete the ratings, please provide written justification for your ratings as you will be required to verbally justify your ratings to me during a scheduled one-on-one session next week. Please consider project-related behaviors only. Please fill-in each of your team members’ names below, including yourself, however remember that your responses will remain anonymous”

If the students received the no accountability (evaluative) TBF, they read the following instructions:

“This form will be used to evaluate you and your team members. This form will be used as a basis for you and your team members’ final grades.

Please fill in a rating for each of your team members beside their code (e.g., M1). Please consider project-related behaviors only. Please fill-in each of your team members’ names below, including yourself, however remember that your responses will remain anonymous.”

After the students completed the TBF, the professors collected them and returned them to the researcher by one of three ways: hand-delivery, mail (e.g., USPS), or email.

## CHAPTER THREE

### RESULTS

#### **Analyses**

Both the lab and field study results were analyzed using 2 x 2 x 2 repeated measures ANOVAs to test the six primary hypotheses. In both studies the independent variables were accountability (accountability, no accountability; between-subjects), purpose (evaluative, developmental; between-subjects), and rater source (self, peer; between-subjects). All of the variables were collected in a team setting but the target unit of analysis was the individual team members' ratings—of themselves (self-rating) and their peers (peer ratings)—on the Team Behaviors Form (TBF). Specifically, in the lab study, a peer rating was the participant's rating of his or her teammate because each team consisted of only two participants. In the field study, a peer rating was the participant's average rating of his or her teammates because each team consisted of three to seven participants. So rater source was a between-subjects variable.

The dependent variable was a unit-weighted composite variable termed the "Big5". This was the average of participants' ratings across the 5 dimensions of the TBF that are considered the primary dimensions of effective team performance: mutual performance monitoring, backup behaviors, adaptability, team orientation, and leadership (Salas et al., 2005). Thus there was a "Big5" variable for self-ratings (Big5Self) and a "Big5" variable for peer ratings (Big5Peer). Although a MANOVA analysis was considered for this study, this analysis would "optimally" (i.e., ordinary least squares) weight the 5 dimension scores to maximize the statistical significance of differences

among the conditions. Because Salas et al. (2005) do not specify relative weights for the 5 factors (and there is no other a-priori theoretical basis for differentially weighting the factors) and because equally weighting the 5 factors is a more conservative test of the hypotheses, a simple unit-weighted composite (average score from the Big 5 questions) was used as the primary dependent variable. Thus, the self-ratings were the average response (from the Big 5 questions) about one's own team behavior and the peer ratings were the average of each of the other team members' average response to the Big 5 questions about oneself.

This primary dependent variable served to measure the degree of leniency or severity in self- and peer ratings in team behaviors, based on the degree of difference between the participants' self-rating of team behaviors and their peers' ratings of those behaviors on each of the Big 5 dimensions of the team performance appraisal. The term "leniency" assumes the differences would be in the direction of participants giving higher self-ratings than peer ratings. However, the analysis detected leniency and severity errors (lower self-ratings than peer ratings of team behaviors).

Prior to the ANOVA analyses outlier analyses were conducted. No obvious outliers were observed. The lowest rating from a team member was typically a 3 (and typically this was only on 1 of the ratings - for example the minimum observed peer rating in the field study was 2.53). And the lowest observed ratings were not in a single team, i.e., there was no team that could be identified as giving unusually low ratings. Likewise there were many ratings (e.g., 14 out of 60 for combined mutual performance monitoring in the field study) at the top of the scale (5) so teams with ratings of 5 were

not unusual (i.e., were not outliers). In order to further check for outliers an analysis was conducted to check for large discrepancies between self-rating and rating by the teammates. This was calculated by using an agreement index; specifically the agreement index is the root-mean-squared differences in their ratings of each other; so perfect agreement would be a 0, while complete disagreement would be a 4. No outliers were observed. The maximum disagreement observed was only 1.25 with a mean disagreement of .64 (sd = .32). A frequency analysis showed that the disagreement index varied uniformly across the observed range of 0.00 to 1.25.

### Lab Study Results

The descriptive statistics for the results of the lab study are shown below in Table 1.

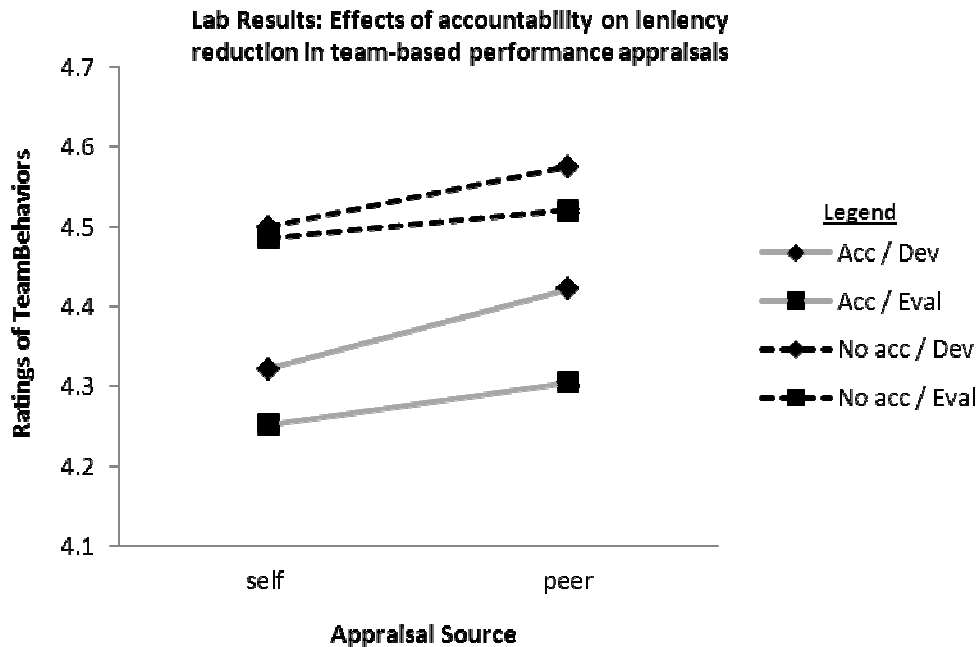
Table 1

#### *Descriptive statistics for the lab study results*

Accountability/Purpose Condition	DV	N	Minimum	Maximum	Mean (SD)
Accountability/Developmental	Big5Self	18	3.60	5.00	4.32 (0.42)
Accountability/Developmental	Big5Peer	18	3.60	5.00	4.42 (0.49)
Accountability/Evaluative	Big5Self	19	3.60	5.00	4.25 (0.49)
Accountability/Evaluative	Big5Peer	19	3.60	5.00	4.31 (0.46)
No accountability/Developmental	Big5Self	24	3.60	5.00	4.50 (0.43)
No accountability/Developmental	Big5Peer	24	3.60	5.00	4.58 (0.38)
No accountability/Evaluative	Big5Self	22	3.00	5.00	4.49 (0.58)
No accountability/Developmental	Big5Peer	24	2.80	5.00	4.52 (0.61)

Based on the results from the ANOVA, none of the hypotheses were clearly supported. No significant 3-way or 2-way interactions were found. No main effect for purpose or source of the appraisal was observed. As seen in Figure 1 below, participants in the no accountability condition did not have significantly higher self-ratings of team behaviors than peer ratings, regardless of the purpose (see hypotheses H1 and H2). Participants in the accountability condition did not have significantly higher self-ratings of team behaviors than peer ratings, regardless of purpose (see hypotheses H3 and H4).

Figure 3. Effects of accountability, purpose, and appraisal source on ratings of team behaviors - Lab study



A marginal main effect for accountability was found ( $df=1$ ,  $F=3.89$ ,  $p = .052$ ; see hypothesis H5). Participants in the accountability condition had somewhat lower ratings of team behaviors, regardless of the purpose. Contrary to Hypothesis 6, participants did not give higher ratings of team behaviors when the appraisals were used for evaluative purposes than when they were used for developmental purposes (see hypothesis H6). A

post-hoc analysis showed no significant difference (within the Accountability condition) for rater source or purpose. Finally, there were no significant interactions between the accountability condition, rater source, and purpose of the appraisal on ratings of team behaviors.

## Field Study Results

The descriptive statistics for the results of the lab study are shown below in Table 2.

Table 2

### *Descriptive statistics for the field study results*

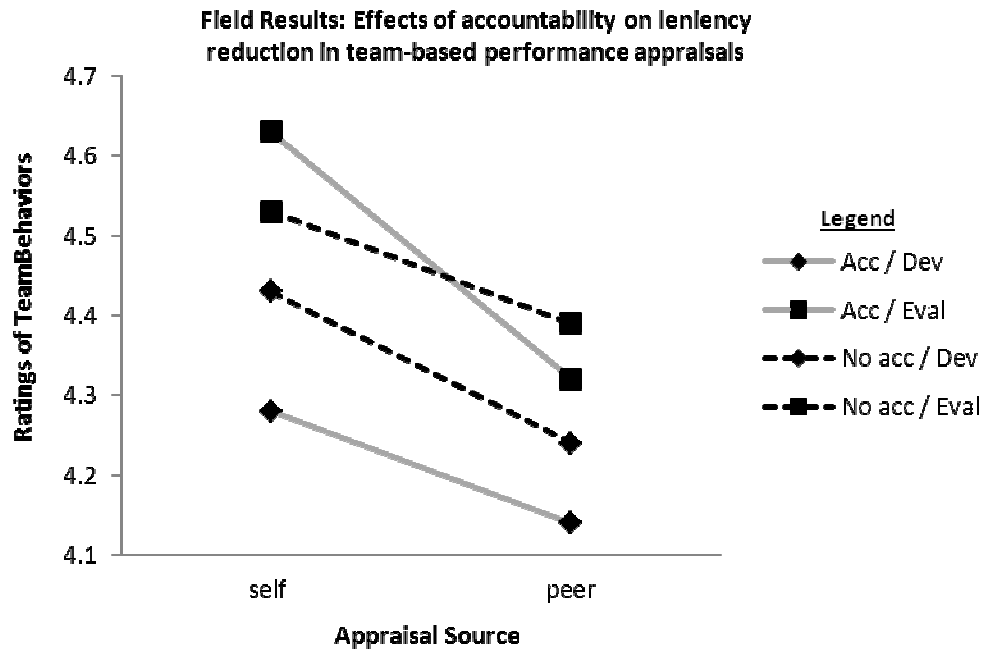
Accountability/Purpose Condition	DV	N	Minimum	Maximum	Mean (SD)
Accountability/Developmental	Big5Self	32	3.00	5.00	4.28 (0.58)
Accountability/Developmental	Big5Peer	38	2.93	5.00	4.14 (0.59)
Accountability/Evaluative	Big5Self	39	3.40	5.00	4.63 (0.39)
Accountability/Evaluative	Big5Peer	40	2.53	5.00	4.32 (0.56)
No accountability/Developmental	Big5Self	40	3.20	5.00	4.43 (0.52)
No accountability/Developmental	Big5Peer	40	2.80	5.00	4.24 (0.51)
No accountability/Evaluative	Big5Self	35	2.80	5.00	4.53 (0.50)
No accountability/Evaluative	Big5Peer	35	2.73	5.00	4.39 (0.65)

Based on the results from the ANOVA, some of the hypotheses were supported by the field study data. No significant 3-way or 2-way interactions were found (see hypotheses H1-H4). As seen in Figure 2 (below) participants had significantly higher self-ratings of team behaviors than peer ratings for both conditions of accountability.



Unlike the lab study, no main effect for accountability was found so hypothesis H5 was not supported.

Figure 4. Effects of accountability, purpose, and appraisal source on ratings of team behaviors - Field study



However, a significant main effect for purpose was observed ( $df=1$ ,  $F=5.90$ ,  $p=.016$ ). These results provide support for hypothesis H6 because participants gave significantly higher ratings of team behaviors when the appraisals were used for evaluative purposes than when they were used for developmental purposes. A significant main effect for appraisal source was observed as well. Self-ratings overall were significantly greater than peer ratings ( $df=1$ ,  $F=24.5$ ,  $p=.0001$ ) As noted there were no significant interactions between the accountability condition, rater source, and purpose of the appraisal on ratings of team behaviors.

## CHAPTER FOUR

### DISCUSSION

The results provided partial evidence for the effectiveness of accountability on leniency reduction in team-based performance appraisals. Specifically, the results did not provide consistent evidence (across both the lab and field studies) that accountability can help reduce leniency when people rate themselves and their peers on team-related job behaviors.

The results for the lab study showed that participants in the accountability conditions provided lower ratings of self- and peer team behaviors than participants in the no accountability condition regardless of purpose showing the potential role that accountability can play in reducing leniency in team performance appraisals. It is possible that, as discussed in the introduction, participants in the accountability conditions assigned lower ratings than participants in the no accountability conditions because the researcher told them they would have to justify their ratings to a local nuclear process control manager. By the researcher telling participants this information, these participants might have assigned ratings that were more reflective of their performance, and might have been less likely to assign overly high ratings of their own and teammate's performance because they believed that the local nuclear process control manager was a credible source. In other words, they might have believed that if they assigned overly high ratings of their own and their teammates' performance, they could not justify these ratings in the face of this manager who would be an expert with the simulator.

Unfortunately, the difference was only marginally significant ( $p = .052$ ) with a fairly

small effect size ( $d = 0.40$ ) so future research will have to continue to test whether accountability can curb leniency reduction on team-based performance appraisals.

Participants might have assigned lower ratings of their own performance in comparison to their peers because they did not feel confident that they mastered the operation of the process control plant. In fact—although the researcher provided participants with clear instructions about how to use the simulator, provided them a chance to practice using the simulator, and asked them if they had any questions (see Appendix A)—when participants in the accountability conditions (i.e., accountability-developmental, accountability evaluative conditions) were asked to elaborate on their self- and peer ratings of team behaviors, many indicated that they had difficulty with the task or were confused about how to operate the process controls. So by assigning lower self-ratings of their performance, participants in the accountability conditions might have believed that they would receive less scrutiny and questioning from the fictitious local nuclear process control manager about their ratings—and about their knowledge of the plant's processes. Although participants in the no accountability condition assigned somewhat higher ratings than participants in the accountability conditions, it is possible that they also assigned lower self than peer ratings of team behaviors because of their unfamiliarity and perceived lack of mastery of the process control plant simulator.

Results for the field study showed a distinctly different pattern than the results from the lab study. The data indicated that participants in the developmental conditions generally assigned significantly lower ratings of team behaviors than participants in the evaluative condition. The finding for a main effect of purpose supports years of research

about the influence that purpose can have on performance ratings in organizational settings. Researchers have found that when performance appraisals are used for evaluative and administrative purposes (e.g., pay raises, promotions, final grades), employees are more likely to inflate their performance ratings in comparison to when appraisals are used for developmental purposes (source). This study adds to that body of research by showing the influence of purpose on the existence of leniency in team-based performance appraisals. After participants completed group assignments—many of which were reflected group assignments that are conducted in organizational settings (e.g., engineering assignment)—it is possible that participants in the evaluative conditions (who read that their ratings would be used as input into their final grades) assigned higher ratings to protect their own and their peers' overall standing in the class compared to their peers in the developmental purpose conditions.

Along with a main effect for purpose, the results also showed a main effect for rater source on participants' ratings of team behaviors. Specifically, the results indicate that across all conditions participants assigned significantly higher self-ratings than peer ratings. This finding supports existing research on the effects of appraisal source on performance, the fact that there is a lack of congruity between self-ratings and other sources on performance appraisals (source). Moreover, this is typically (as was found in the present study) in the direction of higher self-ratings than peer ratings.

In comparison to the lab study, the field study did not provide support for a significant main effect of accountability on leniency reduction in team behaviors of performance. As seen in Figure 3, participants in the accountability-developmental

condition had the lowest self- and peer ratings of team behaviors in comparison to participants in the other three conditions, and this finding is in line with the graphical representation of the hypotheses in Figure 1. These findings show that any intervention to curb leniency in performance appraisals has to be strong enough to blunt the role that purpose plays in the performance appraisals. Although the main effect of purpose was salient in this study, it can still be argued that the accountability manipulation worked to curb leniency tendencies because participants in the accountability-evaluative condition assigned lower peer ratings of team behaviors than participants in the no accountability-evaluative condition. On the other hand, it is possible that these participants assigned lower peer ratings of team behaviors as a self-defense mechanism. Interestingly enough, participants in the accountability-evaluative conditions assigned the highest average self-ratings of team behaviors than participants in the other three conditions. This finding is in contrast to the predicted results about the effects of accountability on leniency reduction on the team behaviors form. When they read that their ratings would be used as a basis for their final grades and that they would have to provide written and verbal justification to their professors, they might have assigned higher self- than peer ratings to set themselves apart from their team members. In fact, participants in the accountability-evaluative condition exhibited the largest average self-peer rating differences (.31) in comparison to participants in the other three conditions. These participants might have been reluctant to provide detailed written justification for their self-ratings because they might have believed that doing so would weaken the professors' perception of them as the top team performer or team player—based on their self-ratings—and hence

negatively influenced their final grades. When participants in the accountability-evaluative condition were asked to elaborate on their ratings, 41% of these respondents did not provide written justification for their self-ratings, and some participants that responded provided inadequate justifications for their self-ratings (e.g., “Overall excellent”). To protect their final grades, these participants might have believed that when they had to meet with their professors to provide verbal justification, they could thwart the discussion of their self-ratings by providing more justification and detail around their ratings of their peers’ team behaviors. In essence, these participants might have planned on using the whole meeting to discuss their teammates’ performance on the group assignments, avoiding the need to make any changes to their self-ratings and thereby protecting their final grades. Thus, it is possible that the TBF should be revised so that its instructions clearly state that participants will have to provide extended written and verbal justification of their own ratings along with their ratings of their peers’ team behaviors on a group task.

The results provide several implications for team-based performance appraisal systems in the workplace. If employees are working in teams and unfamiliar with some of their individual job duties, when asked to rate their own and peers’ job performance, they may possibly provide more severe ratings of their own performance if they are told that they will have to provide verbal justification to a credible source (e.g., supervisor). On the other hand, if employees are working on familiar tasks in teams, rate their own and peers’ performance on these tasks, and are told that their ratings will be used for strictly developmental purposes, these employees will be more likely to inflate their own

ratings in comparison to the ratings they assign peers on the same tasks. Based on these findings, accountability may help curb leniency in team performance appraisals but the strength of accountability depends on how employers use these appraisals within their organizations. If they plan to use the team performance appraisals for evaluative purposes, such as for promotions or pay raises, then they must implement a salient measure of accountability to ensure that their employees are not overestimating their team job performance behaviors. This measure of accountability probably must be communicated verbally by one of the higher-ups (e.g., supervisor, leader) and clearly in writing (on team-based appraisal instructions). If they fail to implement an accountability measure within the team performance appraisal system, then they may be rewarding their employees for mediocre or subpar performance.

Additionally, based on the field study, participants' self-ratings of team behaviors were higher than participants' peer ratings of team behaviors regardless of the purpose of the appraisal form. With that said, across both studies' participants in the accountability-developmental condition had lower self- and peer ratings of team behaviors than the other three conditions. Based on this finding, it is likely that employees should complete team performance appraisals for developmental purposes and accountability should be implemented in this type of appraisal system to reduce employees' tendency to assign high self- and peer ratings of team behaviors. Although participants in an accountability condition may assign higher self-ratings of team behaviors than peer ratings of team behaviors, it is likely that these ratings will not be significantly higher than the ratings

that they assign to their peers' team behaviors (and that their self-ratings will not differ from ratings that their peers assigned to them for their team performance).

### **Limitations and Future Research**

It is important to note the limitations that exist in the current study and to suggest how future research that can address these limitations. A major limitation across both studies was the lack of a standard design for the evaluative purpose conditions across the lab and field study. In the lab study, participants in the evaluative purpose condition were told that they would have to complete the TBF to receive credit for participating in the study; whereas in the field study, participants in the evaluative condition read that their completion of the TBF would be used as part of their own and their peers' final grades. In the lab study, the researcher did not instruct teams to complete the TBF to receive course credit—for themselves and for their teammates—because each team consisted of only two participants, and the researcher could not withhold course credit from participants if their teammates did not complete the TBF. Based on these definitions, some could argue that the lab study manipulation of the evaluative purpose was weaker because the students knew that they would receive credit regardless of whether they completed the TBF. They would also argue that participants in the field study might have inflated their own and their peers' ratings because they thought more negative consequences could possibly result from the TBFs being used for their own and their peers' final grades. So future studies should design evaluative purpose conditions that are standard across different studies and that have more tangible and stronger outcomes (e.g., using team behavior forms to make pay or promotion decisions). Likewise, when the researcher



designed the developmental manipulation across both studies, he did not define if the TBF would be used for improving the rating form or for personal development. Ideally, the goal of the development should be made explicit but it is unlikely that the lack of specificity accounted for the different results between the lab and field study.

Another limitation across both studies is that lack of demographic data collection (e.g., race, gender) for both studies. It is possible that the results might have differed based on demographic variables, so future research should determine whether demographic variables, such as race and gender, are related to people's tendency to assign higher ratings of their performance in comparison to peers' ratings of their performance.

Another potential limitation was the use of the process control plant simulator in the lab study. In this study, many participants indicated that they were unfamiliar with the process control plant simulator task. This lack of familiarity might have caused them to assign harsher self-ratings of their team performance. So, in future studies, researchers should examine the effects of accountability on leniency reduction in laboratory settings by having participants perform a more familiar task, or by having participants perform a task at multiple time points—to increase familiarity. The researcher's use of undergraduate students as participants might have affected the results of the study. These students might not have taken the team performance appraisal exercise seriously and/or lacked experience with completing performance appraisals in comparison to employees. Although some would argue that the results may not be generalizable to work settings because of the undergraduate sample, these students had experience working on team

assignments that were linked to tangible outcomes (e.g., their final grades) similar to employees who complete performance appraisals. Many professors at this university typically design their courses by requiring their students to participate in a team or group assignment. Many of these professors also evaluate the students' performance on these team assignments similar to the way that supervisors evaluate employees on team-based job tasks. Future research should examine whether accountability can curb leniency reduction in employees' self- and peer ratings of team behaviors.

The major limitation of the field study was the manipulation of the accountability condition across classroom settings. As previously mentioned, accountability was operationalized as participants reading instructions to provide written justification on the TBF and reading that they would have to provide verbal justification to their professors during a future meeting. Because the majority of professors—who were assigned to the accountability conditions—did not read the manipulation to their students, these students might not have taken the instructions seriously; or they might have believed that completing the TBF was optional. In fact, one professor told his students that completing the form was optional. Although the effect of accountability was marginal in the lab study, based on Figure 2, participants' ratings of self-and peer ratings of team performance closely aligned with the predicted level of ratings in Figure 1. It could be argued that this alignment resulted from the strength of the accountability manipulation where the experimenter verbally told participants in the accountability conditions to provide written justification on the form and that they would be required to verbally justify their ratings to a superior during a future meeting. Future research should examine

whether the presentation of accountability (written or verbal) affects its power on reducing leniency in performance appraisals.

Another limitation dealt with the design of the field study. In comparison to the lab study, each professor in the field study had a different group assignment for his students. So the results from the field study might have been influenced by the group assignment itself. Therefore the differences in group assignment might have been a potential confound to this study and future research should evaluate the effects of accountability on leniency reduction on team appraisals when participants receive the same group assignment or task.

Also, initially the researcher planned to administer the TBF two times within each classroom during a given semester. However, due to regulatory constraints, the researcher was allowed to administer the form only at the end of the semester. With this one administration, professors administered TBF on the last day of class and gave their students no more than thirty minutes to complete it. So, it is possible that the limited amount of time students had to complete the appraisal, coupled with their unfamiliarity with the form, influenced the outcomes of this study. In comparison to the current study, if students would have received the TBF two times during a given semester, the researcher could have seen how the first administration affected the interpersonal dynamics within each team. If professors did not resolve any interpersonal conflicts between team members, the results from the second administration of the TBF might have exhibited severity effects from team members' retaliation through TBF ratings of their teammates. On the other hand, when students completed the TBF at the end of the

semester, it is possible that these ratings might have been lower and more reflective of their actual performance.

The researcher attempted to overcome the limitation of one administration by instructing the professors to review the “TBF Dimensions Defined” document—which defines the dimensions of team performance that they would use to assess themselves and their peers on their team performance—with their students prior to administering the TBF. After collecting the forms from the professors and asking them about how they administered the forms, they indicated that they simply distributed the “TBF Dimensions Defined” document at the same time as they administered the TBF. Had the professors verbally reviewed the document with the students and if participants were able to complete the TBFs at two different time points, it is possible that the students would have been more familiar on how to complete the TBF.

The field study limitations relate to common performance appraisal administration best practices in organizational settings. In organizational settings, employees should be introduced to appraisal systems well in advance of the time they assess their own performance and before their peers and supervisors assess their performance. This introduction usually entails a meeting between the supervisors and their employees about these employees’ past performance and the supervisor uses this information discuss goals for the current performance year. During this meeting, the supervisor also should explain how employees’ performance will be assessed in terms of performance dimensions and outcomes. Therefore, for future research studies—in order for professors to get more meaningful information from the TBF to inform their

evaluations of their students on group assignments—investigators should work with professors to define an adequate amount of time throughout the semester to introduce the Team Behavior Form in terms of its dimensions and in terms of how to provide written justification of their ratings (if students receive an accountability version of the form).

The researcher might have analyzed the results of the first administration of the TBF to pinpoint weaknesses in how the students completed the form. For instance, many participants in the accountability condition provided insufficient written justification (e.g., leaving sections—where they should have provided written justification for their own and peers' ratings—blank on the TBF) for their self- and peer ratings of team performance. Many of these participants merely stated that they put the necessary amount of effort into the group project assignment and that they worked well with the team; but they did not elaborate on why they assigned certain ratings of team behaviors to themselves and their peers. If the professors administered the TBF during the middle of the semester, the researcher could have pointed out this weakness to the professors and could have instructed them to provide constructive feedback to the students about their insufficient written justifications. After providing this feedback, the researcher could have worked with the professors to provide students with examples on how to write better justifications that provide detailed explanations for why their ratings on the form. Therefore, experimenters should design future research studies by administering the TBF at different time points to allow for a thorough analysis of the students' initial written justifications of their self- and peer ratings on the TBF, and—more important—to

determine if accountability can curb leniency reduction when team-based performance appraisals are used for evaluative or administrative purposes.

In addition to limitations to the design and manipulations of the lab and field studies, there were also limitations with the design of the Team Behaviors Form. Across both studies, the instructions on the accountability version of the form were not expository. They did not explicitly inform participants that they would have to provide written justification for their own and their peers' ratings on each dimension of team behavior. It is possible that the lack of detailed written justification in both studies resulted from the lack of expository instructions on the accountability version of the TBF. Although these explicit instructions were not evident, the accountability version of the form did instruct participants to provide written justification for their ratings. The researcher was reluctant to provide instructions that were too descriptive because he thought the participants would skip reading the instructions under these conditions. The accountability version of the TBF should be revised so that it clearly states that participants will have to provide written justification for their own and peers' ratings on each dimension of the team behaviors form. For the lab study, one major limitation of the TBF is that it did not ask participants to rate their performance on task-related dimensions. On the form, participants rated their own and peers' team behaviors on the process control simulator exercise. It would have been better to include items assessing their own and peers' task performance because the researcher could have compared their subjective ratings of task performance with objective ratings to determine the accuracy of their evaluations. Across both studies, the TBF did not include items where participants

could rate their ability to perform job tasks or their self-efficacy for rating their own and peers' team behaviors on the process control simulator (lab study) or group assignment (field study)—which might have related to the pattern of results for the lab and field studies. Future research should investigate whether leniency in team performance appraisals is tied to employees perceived capability in performing job-related tasks and to their perceived ability to rate their own and others' performance on these tasks.

## **Conclusion**

Although the hypotheses were only partially supported, they add to existing literature by showing the potential effects of accountability on team performance appraisals, and these results have beneficial implications for employers. The primary benefit of implementing accountability into team performance appraisal systems within organizations is that employers will be able to use self- and peer ratings (into these appraisal systems) with greater confidence that they will serve as a reflection of an individual and overall team's actual job performance. Based on research about perceptions of fairness in performance appraisal systems, when employees are allowed to rate themselves and others on team-related job behaviors, it is highly probable that they will perceive the appraisal system as being fair, due to the fact that they play a pivotal role in their own and others' performance evaluations. This will probably lead to employees having more positive perceptions of their organizations, which in turn might lead to an overall increase and improvement in their job performance. In fact, when accountability is incorporated into the team performance appraisal system, it may encourage employees to more critically evaluate their own and their coworkers' team-

related job performance continuously throughout the lifespan of their work teams rather than solely evaluating their job performance once a year. This type of teams-based appraisal system could help supervisors get real-time assessments of their teams' progress in executing particular assignments. This system could help them make corrective personnel actions if team members are reporting poor team performance and behaviors of other team members that is affecting the overall performance and productivity of the team.

Overall, because the hypotheses were partially supported, this study opens up a promising line of research on ways to reduce leniency bias in performance appraisals. Furthermore, this study provides some potential solutions to tackle issues inherent in team performance appraisals. This study can encourage researchers to focus on variables that can directly impact the raters' goals and the stake that they have in the performance appraisal system. This study may also open up avenues for future research that focuses on the effects of individual differences and personality variables on leniency in performance appraisal ratings, and demonstrate how they could influence the effectiveness of variables such as accountability, on leniency reduction in performance appraisals. Also, the overall findings of this study should encourage researchers to examine the longitudinal effects of accountability to determine if this variable can help curb leniency in team-based performance appraisals over time and specifically in those team performance appraisals that are used for administrative purposes.



## APPENDICES

## Appendix A

### Protocol: Process Control Plant Simulator

- 1) Turn on the desktop computers before participants arrive
- 2) Open up operator A setup and type in dem (before participants arrive)
- 3) Open up operator B setup and type in dem (before participants arrive)
- 4) Direct participants to sit at one of two desks labeled “A” and “B” (which are set apart from each other). Have participants read and sign the consent form. Collect the consent forms and then say:

“Hello. The purpose of this lab study is to evaluate the performance of individuals on a process control plant simulation exercise. In this study, you’ll operate a simulated chemical plant, called XPlant, by controlling the fluid that enters and exits the plant. The most important task is to ensure the safety of the plant by working together as a team. Before we begin, do you all know each other? [If they do not] Okay, please briefly introduce yourselves to each other by first name. [After introductions] Before I start describing the XPlant diagram, let’s review the dimensions of effective team performance which you all will use to evaluate each other on this task:  
**\*\*SHOW THEM THE DEFINITIONS SHEET & GO THROUGH IT\*\***  
[After reading the definitions sheet] “Now please go to over to the process control plant simulation.”  
“Now let’s review the diagram on the wall (the same diagrams are presented on your desk). You’ll see this is a basic input-output diagram that shows how that fluid enters and exits the XPlant through each of the five tanks. So the fluid enters the plant through A1 and exits the plant through B2. Operator A controls Subsystem A, tanks A1 and A2 and Operator B controls Subsystem B, tanks B1 and B2. So, essentially Operator A controls the input of the fluid in XPlant and Operator B controls the output of fluid in the XPlant simulator. In addition monitoring your tanks it is also important to monitor the center panel as the fluid enters and exits through this panel to leave the plant.”  
“I’ll open up a session and demonstrate how to use the simulator. Before I walk through each subsystem, remember the overall goal of this exercise is to maintain the safety of the plant. You accomplish this task by monitoring three main indicators for each tank: the fluid level, fluid’s pressure, and fluid temperature. “  
**\*\*When you explain the tanks on one computer monitor, ensure that the other monitor is off. When you explain the center panel, ensure BOTH monitors are off.\*\***

“So looking at the main screen you see the fluid level, pressure, and temperature for Units A1 and A2. For right now, you want to keep the input pump in this tank A1 and output pump in tank B2 constant at 50 units. This means that you want the amount of fluid that enters the plant to equal the amount of fluid that leaves the plant.”

A1: When you click on A1, you see a closer view of A1. For each indicator you'll see a green, yellow, or red bar. Of course a green bar means that an indicator is at the proper setting; yellow means that it is slightly above or below its ideal setting; and a red bar means that the indicator is too far from the ideal setting. Because this is the main input tank, do not change the input pump level. If the temperature is too high, turn on the refrigerant (cooling). If the temperature is too low turn on the heater to make the indicator green. Don't forget to turn off each control after the temp has reached the ideal level. **If the level is too low that means that you are releasing too much fluid from the tank into A2. If this is the case you should reduce the amount of input for tank A2. If the level is too high then you are not releasing enough fluid into A2 so you should increase the level of input into A2.** If the pressure is too low, increase the pressurizer level. If it is too high, ensure that the pressurizer is set to zero and let it level back.

A2: If the temperature is too high, ensure the heater is off and let the temperature settle back to the normal position. If the temperature is too low turn on the heater to bring the temperature back up to a reasonable level. Remember that once the temperature is at the desired level to turn the heat off or reduce it. **Similar to tank A1, if the fluid level in A2 is too high that means you are not releasing enough fluid into the center panel and/or you are inputting too much fluid into this tank from A1. Thus you would need to increase the amount of input into the center panel by adjusting the “pump in” button on the center panel and/or you would need to decrease the input pump (so fluid coming from A1 to A2). If the fluid level is too low you are releasing too much fluid from A2 to the center panel (OR pumping in too much fluid into the center panel from A2) and/or you are not bringing enough fluid from A1 to A2 (the input pump is set too low). So to fix this, reduce the pump in button by turning knob leftwards and/or increase the input pump units.** If the pressure is too high, turn on the vent to reduce the pressure. If the pressure is too low, turn on the pressurizer. When you have brought the pressure level to the desired state make sure to turn off either the vent or the pressurizer.

Center Panel: If the pressure reading is to the left of the blue line that means it is too low. So turn on both pressurizer knobs and wait until the pressure is at the desired state (i.e., meter hits the blue line). Once this occurs turn off both knobs. **For the fluid level the dark green indicator is the desired amount of fluid for the center panel. So if any lights**

above the dark green light are lit, that means that your fluid level is too high, so you are not releasing enough fluid from center panel into B1 and/or you are pumping in too much fluid from A2 into the center tank. In order to reduce the fluid level, the pump out button would need to be increased by turning the knob rightwards (in other words releasing more fluid into B1) and/or the pump in button would need to be decreased by turning the knob leftwards. If level is too low, so if any of the indicator lights below the dark green one are on, that means you are either releasing too much fluid into B1 and/or there is not enough fluid coming into this panel. If you are releasing too much fluid, adjust the pump out button by turning it leftwards and/or adjust the pump in button by turning it rightwards.

So, each tank's fluid level is dependent on other tanks' fluid levels. For instance, the amount of fluid in tank B1 is dependent on the amount of fluid in tanks A1, A2, and B2. The other two indicators do not work this way in that they are exclusive to each tank, so the pressure and temperature in A1 do not depend on the temperature and pressure in the other tanks. Does that make sense?

B1: If the temperature is too high, ensure the heater is off and let the temperature settle back to the normal position. If the temperature is too low turn on the heater to bring the temperature back up to a reasonable level. Remember that once the temperature is at the desired level to turn the heat off or reduce it. **If fluid level is too low, that means you are releasing too much fluid into B2 and/or that you are not pumping out enough fluid from the center panel. To fix this, increase pump out button by turning it rightwards and/or reduce the output pump from B2 to B2. If it is too high it means that you are pumping out too much fluid from the center panel into B1 and/or that you are not releasing enough of fluid into B2. To fix this reduce the pump out level from center panel to B1 by turning knob leftwards and/or increase output pump that releases fluid from B1 to B2.** If the pressure is too high, turn on the vent to reduce the pressure. If the pressure is too low, turn on the pressurizer. When you have brought the pressure level to the desired state make sure to turn off either the vent or the pressurizer

**B2: Because this is the main output tank, do not change the output pump level.** If the temperature is too high, turn on the refrigerant (cooling). If the temperature is too low turn on the heater to make the indicator green. Don't forget to turn off each control after the temp has reached the ideal level. **If the fluid level is too high, you are releasing too much fluid from B1 into B2. If fluid level is too low, that means you are not releasing enough fluid from B1 to B2.** If the pressure is too low, increase the pressurizer level. If it is too high, ensure that the pressurizer is set to zero and let it level back.

So just to recap, you all have to work together as a team to ensure that the fluid is entering and leaving the XPlant at the same rate. Do not change the input pump in tank A1 and do not change the output pump in tank B2 because you want the amount of fluid leaving the plant to equal the amount of fluid entering the plant which is 50 units (percent) for right now. If you change fluid level in one tank it will affect fluid level in another tank. Also, it is very important to monitor the center panel in addition to monitoring your own subsystems.

5) Open Operator B setup and type in “w” (for warmup) “b” for operator B, participant’s number, accountability condition (A or NA), and participant’s purpose condition (D or E). Example code: “wb4nae.”

6) Open Operator A setup and type in “w” (for warmup) “a” for operator A, participant’s number, accountability condition (A or NA), and participant’s purpose condition (D or E). Example code: “wa3nae.”

7) **“Now you all will have a warmup session, where y’all will have to maintain the safety of the plant as 50 units of fluid enters and exits the plant. This session will allow you all to use the different buttons and see how the plant works. Please make sure to collaborate with each other in addition to monitoring your own sub-systems, and make sure to ask me about anything y’all don’t understand. Do y’all have any questions right now? You can begin now.”**

8) End the training simulation, open operator B setup, type in participant’s number, accountability condition (A or NA), and the participant’s purpose condition (D or E). Example code: “b4nae.”

9) Open up the X-plant simulation for operator B

10) Open operator A setup, type in participant’s number, accountability condition (A or NA), and the participant’s purpose condition (D or E). Example code: “a7ad.”

11) Open up the X-plant simulation for operator A

12) **“You all will now have to maintain the safety of the plant as 65 units of fluid enter and exit it and this is real session. Please make sure to collaborate with each other in addition to monitoring your own sub-systems. After the session, you will complete a performance evaluation of yourself and your partner based on the team dimensions we reviewed earlier. Are there any questions? Please let me know if you have any questions. The trial will now begin.”**

13) “Now you are about to complete the Team Behaviors Form ... [read one of the four scenarios (depending on assigned conditions)]:

**NA-D**: “This is a feedback form intended to assess team member behaviors. This form is for developmental purposes only.

Please fill in a rating for yourself and your teammate. Please consider behaviors only related to this process control task. Please fill-in each of your and your teammate’s names below, however remember that your responses will remain anonymous.”

**NA-E**: “This is a feedback form intended to assess team member behaviors. You **MUST** complete this form to receive credit for participation in this study.

Please fill in a rating for yourself and your teammate. Please consider behaviors only related to this process control task. Please fill-in each of your and your teammate’s names below, however remember that your responses will remain anonymous.”

**A-D**: “This is a feedback form intended to assess team member behaviors. This form is for developmental purposes only.

Please fill in a rating for yourself and your teammate. After you complete the ratings, provide written justification on the back of this sheet for your ratings as you will be required to verbally justify your ratings to a local nuclear process control manager during a scheduled one-on-one session next week. Please consider behaviors only related to this process control task. Please fill-in each of your and your teammate’s names below, however remember that your responses will remain anonymous.”

**A-E**: “This is a feedback form intended to assess team member behaviors. You **MUST** complete this form to receive credit for participation in this study.

Please fill in a rating for yourself and your teammate. After you complete the ratings, provide written justification on the back of this sheet for your ratings as you will be required to verbally justify your ratings to a local nuclear process control manager during a scheduled one-on-one session next week. You **MUST** meet with this manager next week to receive credit for participation in this study. Please consider behaviors only related to this process control task. Please fill-in each of your and your teammate’s names below, however remember that your responses will remain anonymous.”

The experimenter will then tell the participants to begin completing the TBF.

14) Once they complete the TBF, give the participants the debriefing sheet and have participants read and sign it. Ask them if they have any questions and allow them to leave the lab.

## Appendix B

### TBF Dimensions Defined Doc

#### **8 Dimensions on the Team Behaviors Form (TBF)**

**Purpose:** The purpose of the Team Behaviors Form (TBF) is to effectively diagnose team processes in order to drive effective team-based performance and outcomes.

1. Team leadership: The capability of a team member to delegate group tasks to other team members, evaluate the level of team performance, foster the enhancement of team members' knowledge skills and abilities, and create a positive environment for the team.
2. Mutual performance monitoring: Team members' ability to create an environment in which they share knowledge of the responsibilities and duties of each other to track individual and team performance.
3. Backup behavior: Team members' proficient knowledge of each other's job duties. This proficient knowledge enables them to foresee the needs of their team members, such as distributing the workload of a team member or team members who are unable to perform their tasks or job duties.
4. Adaptability: A skill that causes team members to be able to shift their tactics and behaviors in response to factors internal and external to the team.
5. Team orientation: The extent to which team members value team goals over individual goals and team members' willingness to work collectively as a unit to accomplish these team goals.
6. Shared mental models: Consist of a common framework in which team members are aware of the relations among team duties and goals and of the interpersonal relationships that team members have with each other.
7. Mutual trust: Team members' perceptions of their ability to execute team tasks and safeguard the psychological, social, and emotional interests of each other.
8. Closed-loop communication: A form of communication that involves the direct transaction between the sender and the recipient regardless of the communication modality (Salas, Sims, & Burke, 2005).

## Appendix C

### Lab TBF: Accountability-Evaluative

Team Number: \_\_\_\_\_

#### **Team Behaviors Form**

This is a feedback form intended to assess team member behaviors. **You MUST complete this form to receive credit for participation in this study.**

Please fill in a rating for yourself and your teammate. After you complete the ratings, provide written justification on the back of this sheet for your ratings as you will be **required** to verbally justify your ratings to a local nuclear process control manager during a scheduled one-on-one session next week. You **MUST** meet with this manager next week to receive credit for participation in this study. Please consider behaviors only related to this process control task. Please fill-in each of your and your teammate's names below, however remember that your responses will remain anonymous:

M1[YOU]: \_\_\_\_\_ M2: \_\_\_\_\_

How would you rate:

1. This team member's awareness of individuals' progress?

Poor		Average		Excellent				
1	-----	2	-----	3	-----	4	-----	5
		M1: _____		M2: _____				

2. This team member's willingness to take on your roles and responsibilities when you needed help?

Poor		Average		Excellent				
1	-----	2	-----	3	-----	4	-----	5
		M1: _____		M2: _____				

3. This team member's adaptation to challenges and unforeseen situations?

Poor		Average		Excellent				
1	-----	2	-----	3	-----	4	-----	5
		M1: _____		M2: _____				



4. This team member's willingness to work with other team members?

Poor                                  Average                                  Excellent

1 ----- 2 ----- 3 ----- 4 ----- 5

M1: _____	M2: _____
-----------	-----------

5. The extent to which this team member is “on the same page” as other team members?

Poor                                  Average                                  Excellent

1 ----- 2 ----- 3 ----- 4 ----- 5

M1: _____	M2: _____
-----------	-----------

6. This team member's communication with other team members?

Poor                                  Average                                  Excellent

1 ----- 2 ----- 3 ----- 4 ----- 5

M1: _____	M2: _____
-----------	-----------

7. Level of mutual trust that this team member has with other team members?

Poor                                  Average                                  Excellent

1 ----- 2 ----- 3 ----- 4 ----- 5

M1: _____	M2: _____
-----------	-----------

8. Do you feel that leadership exists within your team (circle one)?      Yes      No

If so, rate the effectiveness of this team member's leadership abilities in the team?

Poor                                  Average                                  Excellent

1 ----- 2 ----- 3 ----- 4 ----- 5

M1: _____	M2: _____
-----------	-----------

9. Please elaborate on the ratings above **AND** provide any additional suggestions for the improvement of your team's effectiveness.

M1:

M2:

## Lab TBF: Accountability-Developmental

## Team Behaviors Form

Please fill in a rating for yourself and your teammate. After you complete the ratings, provide written justification on the back of this sheet for your ratings as you will be **required** to verbally justify your ratings to a local nuclear process control manager during a scheduled one-on-one session next week. Please consider behaviors only related to this process control task. Please fill-in each of your and your teammate's names below, however remember that your responses will remain anonymous:

How would you rate:

M1: _____	M2: _____
-----------	-----------

M1: _____	M2: _____
-----------	-----------

M1: _____	M2: _____
-----------	-----------

4. This team member's willingness to work with other team members?

Poor                                  Average                                  Excellent

1 ----- 2 ----- 3 ----- 4 ----- 5

M1: _____	M2: _____
-----------	-----------

5. The extent to which this team member is “on the same page” as other team members?

Poor                                  Average                                  Excellent

1 ----- 2 ----- 3 ----- 4 ----- 5

M1: _____	M2: _____
-----------	-----------

6. This team member's communication with other team members?

Poor                                  Average                                  Excellent

1 ----- 2 ----- 3 ----- 4 ----- 5

M1: _____	M2: _____
-----------	-----------

7. Level of mutual trust that this team member has with other team members?

Poor                                  Average                                  Excellent

1 ----- 2 ----- 3 ----- 4 ----- 5

M1: _____	M2: _____
-----------	-----------

8. Do you feel that leadership exists within your team (circle one)?      Yes      No

If so, rate the effectiveness of this team member's leadership abilities in the team?

Poor                                  Average                                  Excellent

1 ----- 2 ----- 3 ----- 4 ----- 5

M1: _____	M2: _____
-----------	-----------

9. Please elaborate on the ratings above **AND** provide any additional suggestions for the improvement of your team's effectiveness.

M1:

M2:





## Appendix F

### Lab TBF: No Accountability-Developmental

Team Number: \_\_\_\_\_

## Team Behaviors Form

This is a feedback form intended to assess team member behaviors. **This form is for developmental purposes only.**

Please fill in a rating for yourself and your teammate. Please consider behaviors only related to this process control task. Please fill-in each of your and your teammate's names below, however remember that your responses will remain anonymous:

M1[YOU]: \_\_\_\_\_ M2: \_\_\_\_\_

How would you rate:

1. This team member's awareness of individuals' progress?

Poor                                  Average                                  Excellent

1 ----- 2 ----- 3 ----- 4 ----- 5

M1: _____	M2: _____
-----------	-----------

2. This team member's willingness to take on your roles and responsibilities when you needed help?

Poor                                  Average                                  Excellent

1 ----- 2 ----- 3 ----- 4 ----- 5

M1: _____	M2: _____
-----------	-----------

3. This team member's adaptation to challenges and unforeseen situations?

Poor                                  Average                                  Excellent

1 ----- 2 ----- 3 ----- 4 ----- 5

M1: _____	M2: _____
-----------	-----------

4. This team member's willingness to work with other team members?

Poor    Average    Excellent

1 ----- 2 ----- 3 ----- 4 ----- 5

M1: _____	M2: _____
-----------	-----------

5. The extent to which this team member is “on the same page” as other team members?

Poor                                  Average                                  Excellent

1 ----- 2 ----- 3 ----- 4 ----- 5

M1: _____	M2: _____
-----------	-----------

6. This team member's communication with other team members?

Poor                                  Average                                  Excellent

1 ----- 2 ----- 3 ----- 4 ----- 5

M1: _____	M2: _____
-----------	-----------

7. Level of mutual trust that this team member has with other team members?

Poor                                  Average                                  Excellent

1 ----- 2 ----- 3 ----- 4 ----- 5

M1: _____	M2: _____
-----------	-----------

8. Do you feel that leadership exists within your team (circle one)?      Yes      No

If so, rate the effectiveness of this team member's leadership abilities in the team?

Poor                                  Average                                  Excellent

1 ----- 2 ----- 3 ----- 4 ----- 5

M1: _____	M2: _____
-----------	-----------

## Appendix G

### Field TBF: Accountability-Evaluative

"1" Team Number: \_\_\_\_\_

#### **Team Behaviors Form**

This is a feedback form intended to assess team member behaviors. **This form will be used to evaluate you and your team members. This form will be used as a basis for your and your team members' final grades.**

Please fill in a rating for each of your team members beside their code (e.g., M1). After you complete the ratings, please provide written justification for your ratings as you will be **required** to verbally justify your ratings to me during a scheduled one-on-one session next week. Please consider project-related behaviors only. Please fill-in each of your team members' names below, **including yourself**, however remember that your responses will remain anonymous:

M1[YOU]: \_\_\_\_\_ M2: \_\_\_\_\_ M3: \_\_\_\_\_  
M4: \_\_\_\_\_ M5: \_\_\_\_\_ M6: \_\_\_\_\_

How would you rate:

1. This team member's awareness of individuals' progress?

Poor Average Excellent  
1 ----- 2 ----- 3 ----- 4 ----- 5

M1: _____	M2: _____	M3: _____	M4: _____	M5: _____	M6: _____
-----------	-----------	-----------	-----------	-----------	-----------

2. This team member's willingness to take on the roles and responsibilities of individuals who need help?

Poor Average Excellent  
1 ----- 2 ----- 3 ----- 4 ----- 5

M1: _____	M2: _____	M3: _____	M4: _____	M5: _____	M6: _____
-----------	-----------	-----------	-----------	-----------	-----------

3. This team member's adaptation to challenges and unforeseen situations?

Poor Average Excellent  
1 ----- 2 ----- 3 ----- 4 ----- 5

M1: _____	M2: _____	M3: _____	M4: _____	M5: _____	M6: _____
-----------	-----------	-----------	-----------	-----------	-----------



4. This team member's willingness to work with other team members?

Poor    Average    Excellent

1 ----- 2 ----- 3 ----- 4 ----- 5

M1: _____	M2: _____	M3: _____	M4: _____	M5: _____	M6: _____
--------------	-----------	-----------	-----------	-----------	-----------

5. The extent to which this team member is “on the same page” as other team members?

Poor                                  Average                                  Excellent

1 ----- 2 ----- 3 ----- 4 ----- 5

M1: _____	M2: _____	M3: _____	M4: _____	M5: _____	M6: _____
-----------	-----------	-----------	-----------	-----------	-----------

6. This team member's communication with other team members?

Poor                                  Average                                  Excellent

1 ----- 2 ----- 3 ----- 4 ----- 5

M1: _____	M2: _____	M3: _____	M4: _____	M5: _____	M6: _____
-----------	-----------	-----------	-----------	-----------	-----------

7. Level of mutual trust that this team member has with other team members?

Poor    Average    Excellent

1 ----- 2 ----- 3 ----- 4 ----- 5

M1: _____	M2: _____	M3: _____	M4: _____	M5: _____	M6: _____
-----------	-----------	-----------	-----------	-----------	-----------

8. Do you feel that leadership exists within your team (circle one)?      Yes      No

If so, rate the effectiveness of this team member's leadership abilities in the team?

Poor                                  Average                                  Excellent

1 ----- 2 ----- 3 ----- 4 ----- 5

M1: _____	M2: _____	M3: _____	M4: _____	M5: _____	M6: _____
-----------	-----------	-----------	-----------	-----------	-----------

9. Please elaborate on the ratings above **AND** provide any additional suggestions for the improvement of your team's effectiveness.

M1:

M2:

M3:

M4:

M5:

M6:

## Appendix H

### Field TBF: Accountability-Developmental

"1" Team Number: \_\_\_\_\_

#### **Team Behaviors Form**

This is a feedback form intended to assess team member behaviors. **This form will NOT be used to evaluate or grade you or your team members, it is for developmental purposes only.**

Please fill in a rating for each of your team members beside their code (e.g., M1). After you complete the ratings, please provide written justification for your ratings as you will be **required** to verbally justify your ratings to me during a scheduled one-on-one session next week. Please consider project-related behaviors only. Please fill-in each of your team members' names below, **including yourself**, however remember that your responses will remain anonymous:

M1[YOU]: \_\_\_\_\_ M2: \_\_\_\_\_ M3: \_\_\_\_\_  
M4: \_\_\_\_\_ M5: \_\_\_\_\_ M6: \_\_\_\_\_

How would you rate:

1. This team member's awareness of individuals' progress?

Poor Average Excellent  
1 ----- 2 ----- 3 ----- 4 ----- 5

M1: _____	M2: _____	M3: _____	M4: _____	M5: _____	M6: _____
-----------	-----------	-----------	-----------	-----------	-----------

2. This team member's willingness to take on the roles and responsibilities of individuals who need help?

Poor Average Excellent  
1 ----- 2 ----- 3 ----- 4 ----- 5

M1: _____	M2: _____	M3: _____	M4: _____	M5: _____	M6: _____
-----------	-----------	-----------	-----------	-----------	-----------

3. This team member's adaptation to challenges and unforeseen situations?

Poor Average Excellent  
1 ----- 2 ----- 3 ----- 4 ----- 5

M1: _____	M2: _____	M3: _____	M4: _____	M5: _____	M6: _____
-----------	-----------	-----------	-----------	-----------	-----------

4. This team member's willingness to work with other team members?

Poor                                  Average                                  Excellent

1 ----- 2 ----- 3 ----- 4 ----- 5

M1: ____	M2: ____	M3: ____	M4: ____	M5: ____	M6: ____
----------	----------	----------	----------	----------	----------

5. The extent to which this team member is “on the same page” as other team members?

Poor                                  Average                                  Excellent

1 ----- 2 ----- 3 ----- 4 ----- 5

M1: _____	M2: _____	M3: _____	M4: _____	M5: _____	M6: _____
-----------	-----------	-----------	-----------	-----------	-----------

6. This team member's communication with other team members?

Poor                                  Average                                  Excellent

1 ----- 2 ----- 3 ----- 4 ----- 5

M1: _____	M2: _____	M3: _____	M4: _____	M5: _____	M6: _____
-----------	-----------	-----------	-----------	-----------	-----------

7. Level of mutual trust that this team member has with other team members?

Poor                                  Average                                  Excellent

1 ----- 2 ----- 3 ----- 4 ----- 5

M1: _____	M2: _____	M3: _____	M4: _____	M5: _____	M6: _____
-----------	-----------	-----------	-----------	-----------	-----------

8. Do you feel that leadership exists within your team (circle one)?      Yes      No

If so, rate the effectiveness of this team member's leadership abilities in the team?

Poor                                  Average                                  Excellent

1 ----- 2 ----- 3 ----- 4 ----- 5

M1: _____	M2: _____	M3: _____	M4: _____	M5: _____	M6: _____
-----------	-----------	-----------	-----------	-----------	-----------

9. Please elaborate on the ratings above **AND** provide any additional suggestions for the improvement of your team's effectiveness.

M1:

M2:

M3:

M4:

M5:

M6:

## Appendix I

### Field TBF: No Accountability-Evaluative

"2" Team Number: \_\_\_\_\_

#### **Team Behaviors Form**

This is a feedback form intended to assess team member behaviors. **This form will be used to evaluate you and your team members. This form will be used as a basis for your and your team members' final grades.**

Please fill in a rating for each of your team members beside their code (e.g., M1). Please consider project-related behaviors only. Please fill-in each of your team members' names below, **including yourself**, however remember that your responses will remain anonymous:

M1[YOU]: \_\_\_\_\_ M2: \_\_\_\_\_ M3: \_\_\_\_\_  
M4: \_\_\_\_\_ M5: \_\_\_\_\_ M6: \_\_\_\_\_

How would you rate:

1. This team member's awareness of individuals' progress?

Poor Average Excellent  
1 ----- 2 ----- 3 ----- 4 ----- 5

M1: _____	M2: _____	M3: _____	M4: _____	M5: _____	M6: _____
-----------	-----------	-----------	-----------	-----------	-----------

2. This team member's willingness to take on the roles and responsibilities of individuals who need help?

Poor Average Excellent  
1 ----- 2 ----- 3 ----- 4 ----- 5

M1: _____	M2: _____	M3: _____	M4: _____	M5: _____	M6: _____
-----------	-----------	-----------	-----------	-----------	-----------

3. This team member's adaptation to challenges and unforeseen situations?

Poor Average Excellent  
1 ----- 2 ----- 3 ----- 4 ----- 5

M1: _____	M2: _____	M3: _____	M4: _____	M5: _____	M6: _____
-----------	-----------	-----------	-----------	-----------	-----------

4. This team member's willingness to work with other team members?

Poor                                  Average                                  Excellent

1 ----- 2 ----- 3 ----- 4 ----- 5

M1: ____	M2: ____	M3: ____	M4: ____	M5: ____	M6: ____
----------	----------	----------	----------	----------	----------

5. The extent to which this team member is “on the same page” as other team members?

Poor                                  Average                                  Excellent

1 ----- 2 ----- 3 ----- 4 ----- 5

M1: _____	M2: _____	M3: _____	M4: _____	M5: _____	M6: _____
-----------	-----------	-----------	-----------	-----------	-----------

6. This team member's communication with other team members?

Poor                                  Average                                  Excellent

1 ----- 2 ----- 3 ----- 4 ----- 5

M1: ____	M2: ____	M3: ____	M4: ____	M5: ____	M6: ____
----------	----------	----------	----------	----------	----------

7. Level of mutual trust that this team member has with other team members?

Poor                                  Average                                  Excellent

1 ----- 2 ----- 3 ----- 4 ----- 5

M1: _____	M2: _____	M3: _____	M4: _____	M5: _____	M6: _____
-----------	-----------	-----------	-----------	-----------	-----------

8. Do you feel that leadership exists within your team (circle one)?      Yes      No

If so, rate the effectiveness of this team member's leadership abilities in the team?

Poor                                  Average                                  Excellent

1 ----- 2 ----- 3 ----- 4 ----- 5

M1: _____	M2: _____	M3: _____	M4: _____	M5: _____	M6: _____
-----------	-----------	-----------	-----------	-----------	-----------

## Appendix J

### Field TBF: No Accountability-Developmental

"2" Team Number: \_\_\_\_\_

## Team Behaviors Form

This is a feedback form intended to assess team member behaviors. **This form will NOT be used to evaluate or grade you or your team members, it is for developmental purposes only.**

Please fill in a rating for each of your team members beside their code (e.g., M1). Please consider project-related behaviors only. Please fill-in each of your team members' names below, **including yourself**, however remember that your responses will remain anonymous:

M1[you]: \_\_\_\_\_ M2: \_\_\_\_\_ M3: \_\_\_\_\_  
M4: \_\_\_\_\_ M5: \_\_\_\_\_ M6: \_\_\_\_\_

How would you rate:

1. This team member's awareness of individuals' progress?

Poor    Average    Excellent

1 ----- 2 ----- 3 ----- 4 ----- 5

M1: _____	M2: _____	M3: _____	M4: _____	M5: _____	M6: _____
-----------	-----------	-----------	-----------	-----------	-----------

2. This team member's willingness to take on the roles and responsibilities of individuals who need help?

Poor                                  Average                                  Excellent

1 ----- 2 ----- 3 ----- 4 ----- 5

M1: _____	M2: _____	M3: _____	M4: _____	M5: _____	M6: _____
-----------	-----------	-----------	-----------	-----------	-----------

3. This team member's adaptation to challenges and unforeseen situations?

Poor    Average    Excellent

1 ----- 2 ----- 3 ----- 4 ----- 5

M1: _____	M2: _____	M3: _____	M4: _____	M5: _____	M6: _____
-----------	-----------	-----------	-----------	-----------	-----------



4. This team member's willingness to work with other team members?

Poor                                  Average                                  Excellent

1 ----- 2 ----- 3 ----- 4 ----- 5

M1: _____	M2: _____	M3: _____	M4: _____	M5: _____	M6: _____
-----------	-----------	-----------	-----------	-----------	-----------

5. The extent to which this team member is “on the same page” as other team members?

Poor                                  Average                                  Excellent

1 ----- 2 ----- 3 ----- 4 ----- 5

M1: _____	M2: _____	M3: _____	M4: _____	M5: _____	M6: _____
-----------	-----------	-----------	-----------	-----------	-----------

6. This team member's communication with other team members?

Poor                                  Average                                  Excellent

1 ----- 2 ----- 3 ----- 4 ----- 5

M1: _____	M2: _____	M3: _____	M4: _____	M5: _____	M6: _____
-----------	-----------	-----------	-----------	-----------	-----------

7. Level of mutual trust that this team member has with other team members?

Poor                                  Average                                  Excellent

1 ----- 2 ----- 3 ----- 4 ----- 5

M1: _____	M2: _____	M3: _____	M4: _____	M5: _____	M6: _____
-----------	-----------	-----------	-----------	-----------	-----------

8. Do you feel that leadership exists within your team (circle one)?      Yes      No

If so, rate the effectiveness of this team member's leadership abilities in the team?

Poor                                  Average                                  Excellent

1 ----- 2 ----- 3 ----- 4 ----- 5

M1: _____	M2: _____	M3: _____	M4: _____	M5: _____	M6: _____
-----------	-----------	-----------	-----------	-----------	-----------

## Appendix K

### Field Study IRB Approval Form

**Fred Switzer**

---

**From:** Nalinee Patin  
**Sent:** Tuesday, March 08, 2011 3:52 PM  
**To:** Fred Switzer; Brett Smith  
**Subject:** Not Human Subjects Research Determination for IRB2011-075

Dear Dr. Switzer and Mr. Smith,

The Clemson University Office of Research Compliance (ORC) has determined that your project entitled "**Performance Assessment in Virtual and Physical Classroom Settings**" is not subject to IRB review.

Per Mr. Smith, at this time, the project will not involve the use of "identifiable private information" about living individuals; the data will be de-identified by the instructors before it is shared with the research team. Therefore, use of these data does not involve human subjects as defined in the federal regulations governing the protection of human subjects in research [45 CFR 46.102(f)], and IRB review is not required.

Please contact this office again if there are any changes to this project that might bring it under the purview of the IRB. It is the responsibility of the ORC to determine whether any specific project falls within the definition of research with human subjects, as provided by federal regulations and institutional policy.

Good luck with your project, and let us know if you have any further questions.

Regards,  
Nalinee

*Nalinee D. Patin*  
IRB Coordinator  
Clemson University  
Office of Research Compliance  
Institutional Review Board (IRB)  
Voice: (864) 656-0636  
Fax: (864) 656-4475  
E-mail: [npatin@clemson.edu](mailto:npatin@clemson.edu)  
Web site: <http://www.clemson.edu/research/compliance/irb/>  
IRB E-mail: [irb@clemson.edu](mailto:irb@clemson.edu)

**Confidentiality Notice:** This message is intended for the use of the individual to which it is addressed and may contain information that is confidential. If the reader of this message is not the intended recipient, you are hereby notified that any dissemination, distribution, or copying of this communication is strictly prohibited. If you receive this communication in error, please notify us by reply mail and delete the original message.

## References

- Albright, M. D., & Levy, P. E. (1995). The effects of source credibility and performance rating discrepancy on reactions to multiple raters. *Journal of Applied Social Psychology, 25*(7), 577-600.
- Alicke, M. D., & Sedikides, C. (2009). Self-enhancement and self-protection: What they are and what they do. *European Review of Social Psychology, 20*, 1-48.
- Anderson, C., Ames, D. R., & Gosling, S. D. (2008). Punishing hubris: The perils of overestimating one's status in a group. *Personality and Social Psychology Bulletin, 34*(1), 90-101.
- Antonioni, D., & Park, H. (2001). The relationship between rater affect and three sources of 360-degree feedback ratings. *Journal of Management, 27*(4), 479-495.
- Arthur, W., Edwards, B. D., Bell, S. T., Villado, A. J., & Bennett, W. (2005). Team task analysis: Identifying tasks and jobs that are team based. *Human Factors, 47*(3), 654-669.
- Arvey, R. D., & Murphy, K. R. (1998). Performance evaluation in work settings. *Annual Review of Psychology, 49*, 141-168.
- Baltes, B. B., & Parker, C. P. (2000). Reducing the effects of performance expectations on behavioral ratings. *Organizational Behavior and Human Decision Processes, 82*(2), 237-267.
- Bamberger, P. A., Erev, I., Kimmel, M., & Oref-Chen, T. (2005). Peer assessment, individual performance, and contribution to group processes: The impact of rater anonymity. *Group & Organization Management, 30*(4), 344-377.
- Barfield, R. L. (2002). Students' perceptions of being graded as a group in the college classroom: Relations among students' age, employment, and perceived group, satisfaction. *Perceptual and Motor Skills, 95*(3), 1267-1277.
- Bauer, C. C., & Baltes, B. B. (2002). Reducing the effects of gender stereotypes on performance evaluations. *Sex Roles, 47*(9-10), 465-476.
- Beaulieu, R. P. (2005). Race of ratee and anonymity of rater: A study comparing students with practitioners as performance appraisers. *Psychological Reports, 97*(2), 389-399.

- Becker, G. A., & Miller, C. E. (2002). Examining contrast effects in performance appraisals: Using appropriate controls and assessing accuracy. *Journal of Psychology, 136*(6), 667-683.
- Bergee, M. J., & Cecconi-Roberts, L. (2002). Effects of small-group peer interaction on self-evaluation of music performance. *Journal of Research in Music Education, 50*(3), 256-268.
- Bernardin, H. J., Cooke, D. K., & Villanova, P. (2000). Conscientiousness and agreeableness as predictors of rating leniency. *Journal of Applied Psychology, 85*(2), 232-236.
- Bernardin, H., Tyler, C. L., & Villanova, P. (2009). Rating level and accuracy as a function of rater personality. *International Journal of Selection and Assessment, 17*(3), 300-310.
- Bettenhausen, K. L., & Fedor, D. B. (1997). Peer and upward appraisals: A comparison of their benefits and problems. *Group & Organization Management, 22*(2), 236-263.
- Bretz, Jr., R. D., Milkovich, G. T., & Read, W. (1992). The current state of performance appraisal research and practice: Concerns, directions, and implications. *Journal of Management, 18*, 321-352.
- Borman, W. C., White, L. A., & Dorsey, D. W. (1995). Effects of ratee task-performance and interpersonal factors on supervisor and peer performance ratings. *Journal of Applied Psychology, 80*(1), 168-177.
- Boswell, W. R., & Boudreau, J. W. (2002). Separating the developmental and evaluative performance appraisal uses. *Journal of Business and Psychology, 16*(3), 391-412.
- Brannick, M. T., Prince, A., Prince, C., & Salas, E. (1995). The measurement of team process. *Human Factors, 37*(3), 641-651.
- Brutus, S., & Fecteau, J. (2003). Short, simple, and specific: The influence of item design characteristics in multi-source assessment contexts. *International Journal of Selection and Assessment, 11*(4), 313-325.
- Brutus, S., Fletcher, C., & Baldry, C. (2009). The influence of self-construal on rater self-efficacy in performance appraisal. *International Journal of Human Resource Management, 20*(9), 1999-2011.

- Campbell, D. J., Campbell, K. M., & Chia, H. B. (1998). Merit pay, performance appraisal, and individual motivation: An analysis and alternative. *Human Resource Management, 37*(2), 131-146.
- Çiçek, M. C., Köksal, G., & Özdemirel, N. E. (2005). A team performance measurement model for continuous improvement. *Total Quality Management & Business Excellence, 16*(3), 331-349.
- Colella, A., DeNisi, A. S., & Varma, A. (1997). Appraising the performance of employees with disabilities: A review and model. *Human Resource Management Review, 7*(1), 27-53.
- Colella, A., DeNisi, A., & Varma, A. (1998). The impact of ratee's disability on performance judgments and choice as partner: The role of disability job fit stereotypes and interdependence of rewards. *Journal of Applied Psychology, 83*(1), 102-111.
- Colella, A., & Varma, A. (1999). Disability-job fit stereotypes and the evaluation of persons with disabilities at work. *Journal of Occupational Rehabilitation, 9*(2), 79-95.
- Collins, D. R., & Stukas, A. A. (2008). Narcissism and self-presentation: The moderating effects of accountability and contingencies of self-worth. *Journal of Research in Personality, 42*(6), 1629-1634.
- Cook, D. A., Dupras, D. M., Beckman, T. J., Thomas, K. G., & Pankratz, V. S. (2009). Effect of Rater Training on Reliability and Accuracy of Mini-CEX Scores: A Randomized, Controlled Trial. *Journal of General Internal Medicine, 24*(1), 74-79.
- Cook, D. A., & Beckman, T. J. (2009). Does scale length matter? A comparison of nine-versus five-point rating scales for the mini-CEX. *Advances in Health Sciences Education, 14*(5), 655-664.
- Cunningham, G. B., & Dixon, M. A. (2003). New perspectives concerning performance appraisals of intercollegiate coaches. *Quest, 55*(2), 177-192.
- Curtis, A. B., Harvey, R. D., & Ravden, D. (2005). Sources of political distortions in performance appraisals: Appraisal purpose and rater accountability. *Group & Organization Management, 30*(1), 42-60.

- Daradoumis, T., Martinez-Mones, A., & Xhafa, F. (2004). An integrated approach for analysing and assessing the performance of virtual learning groups. *Groupware: Design, Implementation, and Use, Proceedings* (from *Lecture Notes in Computer Science*), 3198, 289-304.
- DeVoe, S. E., & Iyengar, S. S. (2004). Managers' theories of subordinates: A cross-cultural examination of manager perceptions of motivation and appraisal of performance. *Organizational Behavior and Human Decision Processes*, 93(1), 47-61.
- Diefendorff, J. M., Silverman, S. B., & Greguras, G. J. (2005). Measurement equivalence and multisource ratings for non-managerial positions: Recommendations for research and practice. *Journal of Business and Psychology*, 399-425.
- Dierdorff, E. C., & Surface, E. A. (2007). Placing peer ratings in context: Systematic influences beyond ratee performance. *Personnel Psychology*, 60(1), 93-126
- Dominick, P. G., Reilly, R. R., & McGourty, J. W. (1997). The effects of peer feedback on team member behavior. *Group & Organization Management*, 22(4), 508-520.
- Drexler, J. A., Beehr, T. A., & Stetz, T. A. (2001). Peer appraisals: Differentiation of individual performance on group tasks. *Human Resource Management*, 40(4), 333-345.
- Druskat, V. U., & Wolff, S. B. (1999). Effects and timing of developmental peer appraisals in self-managing work groups. *Journal of Applied Psychology*, 84(1), 58-74.
- Duarte, N. T., Goodson, J. R., & Klich, N. R. (1993). How do I like thee: Let me appraise the ways. *Journal of Organizational Behavior*, 14(3), 239-249.
- Dunning, D. G., Durham, T. M., Aksu, M. N., & Lange, B. M. (2008). The state of the art in evaluating the performance of assistant and associate deans as seen by deans and assistant and associate deans. *Journal of Dental Education*, 72(4), 458-471.
- Edwards, M. R., & Sproull, J. R. (1985). Making Performance Appraisals Perform: The Use Of Team Evaluation. *Personnel*, 62(3), 28-32.
- Ellis, A. P. J., Bell, B. S., Ployhart, R. E., Hollenbeck, J. R., & Ilgen, D. R. (2005). An evaluation of generic teamwork skills training with action teams: Effects on cognitive and skill-based outcomes. *Personnel Psychology*, 58(3), 641-672.

- Ellis, A. P. J., Ilgen, D. R., & Hollenbeck, J. R. (2006). The effects of team leader race on performance evaluations: An attributional perspective. *Small Group Research*, 37(3), 295-322.
- Emery, Y. (2004). Rewarding civil service performance through team bonuses: Findings, analysis and recommendations. *International Review of Administrative Sciences*, 70(1), 157-168.
- English, A., Griffith, R. L., & Steelman, L. A. (2004). Team performance: The effect of team conscientiousness and task type. *Small Group Research*, 35(6), 643-665.
- Erez, A., Lepine, J. A., & Elms, H. (2002). Effects of rotated leadership and peer evaluation on the functioning and effectiveness of self-managed teams: A quasi-experiment. *Personnel Psychology*, 55(4), 929-948.
- Evans, A. W., Leeson, R. M. A., & Petrie, A. (2007). Reliability of peer and self-assessment scores compared with trainers' scores following third molar surgery. *Medical Education*, 41(9), 866-872.
- Evans, R., Elwyn, G. & Edwards, A. (2004). Review of instruments for peer assessment on physicians. *British Medical Journal*, 328(7450), 1240-1243.
- Facteau, J. D., & Craig, S. B. (2001). Are performance appraisal ratings from different rating sources comparable? *Journal of Applied Psychology*, 86(2), 215-227.
- Falchikov, N. (1995). Peer feedback marking: Developing peer assessment. *Innovations in Education and Training International*, 32(2), 175-187.
- Faure, C. (2009). Attribution biases in the evaluation of new product development team members. *Journal of Product Innovation Management*, 26(4), 407-423.
- Fedor, D. B., Bettenhausen, K. L., & Davis, W. (1999). Peer reviews: Employees' dual roles as raters and recipients. *Group and Organization Management*, 24(1), 92-120.
- Findley, H. M., Giles, W. F., & Mossholder, K. W. (2000). Performance appraisal process and system facets: Relationships with contextual performance. *Journal of Applied Psychology*, 85(4), 634-640.
- Fletcher, C. (2001). Performance appraisal and management: The developing research agenda. *Journal of Occupational and Organizational Psychology*, 74, 473-487.

- Fletcher, T. D., & Major, D. A. (2006). The effects of communication modality on performance and self-ratings of teamwork components. *Journal of computer-mediated Communication*, 11(2), 557-576.
- Forsterling, F., & Morgenstern, M. (2002). Accuracy of self-assessment and task performance: Does it pay to know the truth? *Journal of Educational Psychology*, 94(3), 576-585.
- Forsyth, D. R., & Kelley, K. N. (1994). Attribution in groups: Estimations of personal contributions to collective endeavors. *Small Group Research*, 25(3), 367-383.
- Fox, S., Bizman, A., & Garti, A. (2005). Is distributional appraisal more effective than traditional performance appraisal? *European Journal of Psychological Assessment* 21(3), 165-172.
- Frink, D. D., & Ferris, G. R. (1998). Accountability, impression management, and goal setting in the performance evaluation process. *Human Relations*, 51(10), 1259-1283.
- Furnham, A., & Stringfield, P. (1998). Congruence in job-performance ratings: A study of 360 degrees feedback examining self, managers, peers, and consultant ratings. *Human Relations*, 51(4), 517-530.
- Galbraith, R. M., Hawkins, R. E., & Holmboe, E. S. (2008). Making self-assessment more effective. *Journal of Continuing Education in The Health Professions*, 28(1), 20-24.
- Gerard, R. J. (2002). Peer evaluation as a predictor of future success. Ph.D. dissertation, The Pennsylvania State University, US -- Pennsylvania. Retrieved March 17, 2010, from Dissertations & Theses: Full Text. (Publication No. AAI3051652).
- Gilbert, G. R., & Whiteside, C. W. (1988). Performance appraisal and followership: An analysis of the officer on the boss-subordinate team. *Journal of Police Science & Administration*, 16(1), 39-43.
- Goffin, R. D., Jelley, R. B., Powell, D. M., & Johnsnton, N. G. (2009). Taking advantage of social comparisons in performance appraisal: The relative percentile method. *Human Resource Management*, 48(2), 251-268.
- Golden, T. D., Barnes-Farrell, J. L., & Mascharka, P. B. (2009). Implications of Virtual Management for Subordinate Performance Appraisals: A Pair of Simulation Studies. *Journal of Applied Social Psychology*, 39(7), 1589-1608.



- Gomez, C., Kirmkan, B. L., & Shapiro, D. L. (2000). The impact of collectivism and in-group/out-group membership on the evaluation generosity of team members. *Academy of Management Journal*, 43(6), 1097-1106.
- Greguras, G. J., Robie, C., Born, M. P., & Koenigs, R. J. (2007). A social relations analysis of team performance ratings. *International Journal of Selection and Assessment*, 15(4), 434-448.
- Greguras, G. J., Robie, C., Schleicher, D. J., & Goff, M. (2003). A field study of rating on the quality of multisource ratings. *Personnel Psychology*, 56(1), 1-21.
- Hader, R., Sorensen, E. R., Edelson, W., Bliss-Holtz, J. (1999). Developing a registered nurse performance appraisal tool. *Journal of Nursing Administration*, 29(9), 26-32.
- Hannum, K. M. (2007). Measurement Equivalence of 360 degrees-Assessment Data: Are different raters rating the same constructs? *International Journal of Selection and Assessment*, 15(3), 293-301.
- Hardin, A. M., Fuller, M. A., & Valacich, J. S. (2006). Measuring group efficacy in virtual teams: New questions in old debate. *Small Group Research*, 37(1), 65-85.
- Harris, M. M., Smith, D. E., and Champagne, D. (1995). A field study of performance appraisal purpose: Research versus administrative based ratings. *Personnel Psychology*, 48, 151-160.
- Hartmann, F., & Slapnicar, S. (2009). How formal performance evaluation affects trust between superior and subordinate managers. *Accounting Organizations and Society*, 34(6-7), 722-737.
- Heidemeier, H., & Moser, K. (2009). Self-Other Agreement in Job Performance Ratings: A Meta-Analytic Test of a Process Model. *Journal of Applied Psychology*, 94(2), 353-370.
- Heslin, P. A., Latham, G. P., & VandeWalle, D. (2005). The effect of implicit person theory on performance appraisals. *Journal of Applied Psychology*, 90(5), 842-856.
- Hoffman, B. J., Lance, C. E., Bynum, B., & Gentry, W. A. (2010). Rater source effects are alive and well after all. *Personnel Psychology*, 63(1), 119-151.
- Hoffman, B. J., & Woehr, D. J. (2009). Disentangling the meaning of multisource performance rating source and dimension factors. *Personnel Psychology*, 62(4), 735-765.

- Hoffman, J. R., & Rogelberg, S. G. (2001). All together now? College students' preferred project group grading procedures. *Group Dynamics*, 5(1), 33-40.
- Hughes, C., Toohey, S., & Velan, G. (2008). eMed Teamwork: A self-moderating system to gather peer feedback for developing and assessing teamwork skills. *Medical Teacher*, 30(1), 5-9.
- Janz, T., & Harel, G. H. (1993). Performance appraisal for TQM: A team approach. *Total Quality Management*, 4(3), 275-281.
- Jawahar, I. M. (2001). Attitudes, self-monitoring, and appraisal behaviors. *Journal of Applied Psychology*, 86(5), 875-883.
- Jawahar, I. M. (2005). Do raters consider the influence of situational factors on observed performance when evaluating performance? Evidence from three experiments. *Group & Organization Management*, 30(1), 6-41.
- Jawahar, I. M., & Williams, C. R. (1997). Where all the children are above average: The performance appraisal purpose effect. *Personnel Psychology*, 50(4), 905-925.
- Jones, M. C., & Harrison, A. W. (1996). IS project team performance: An empirical assessment. *Information & Management*, 31, 57-65.
- Jordan, J. L., & Jordan, D. N. (1993). Satisfaction with performance appraisal ratings. *Psychological reports*, 72(3), 1222-1222.
- Jordan, J. L., & Nasis, D. B. (1992). Preferences for performance-appraisal based on method used, type of rater, and purpose of evaluation. *Psychological reports*, 70(3), 963-969.
- Kam, S. M. (1996). Self-ratings of performance as reflections of narcissism, impression management, and social comparison. Ph.D. dissertation, University of California at Berkeley, United States -- California. Retrieved March 17, 2010, from Dissertations & Theses: Full Text. (Publication No. AAM9621208)
- Kane, J. S., Bernardin, H. J., Villanova, P., & Peyrefitte, J. (1995). Stability of rater leniency: 3 studies. *Academy of Management Journal*, 38(4), 1036-1051.
- Karakowsky L., & McBey, K. (2001). Do my contributions matter? The influence of imputed expertise on member involvement, and self-evaluations in the work group. *Group & Organization Management*, 26(1), 70-92.
- Kasten, R., & Nevo, B. (2008). Exploring the relationship between interrater correlations and validity of peer ratings. *Human Performance*, 21(2), 180-197.

- Keefe, M., Glancey, J., & Cloud, N. (2007). Assessing student team performance on industry sponsored design projects. *Journal of Mechanical Design*, 129(7), 692-700.
- Klein, W. M. P., & Buckingham, J. T. (2002). Self-other biases in judgments of ambiguous performance and corresponding ability. *Psychologica Belgica*, 42(1-2), 43-64.
- Kline, T. J. B. (2001). Team members' and supervisors' ratings of team performance: A case of inconsistency. *Psychological Reports*, 88, 1015-1022.
- Kline, T. J. B., & Sulksy, L. M. (2009). Measurement and assessment issues in performance appraisals. *Canadian Psychology*, 50(3), 161-171.
- Kudisch, J. D., Fortunato, V.J., & Smith, A. F. R. (2006). Contextual and individual factors predicting individuals' desire to provide upward feedback. *Group & Organization Management*, 31(4), 503-529.
- Kurt, A. & Paulhus, D. L. (2008). Moderators of the adaptiveness of self-enhancement: Operationalization, motivational domain, adjustment facet, and evaluator. *Journal of Research In Personality*, 42(4), 839-853.
- Kurtzberg, T. R., Naquin, C. E., & Belkin, L. Y. (2005). Electronic performance appraisals: The effects of e-mail communication on peer ratings in actual and simulated environments. *Organizational Behavior and Human Decision Processes*, 98(2), 216-226.
- Kuvaas, B. (2007). Different relationships between perceptions of developmental performance appraisal and work performance. *Personnel Review*, 36(3), 378-397.
- Kuvaas, B. (2006). Performance appraisal satisfaction and employee outcomes: Mediating and moderating roles of work motivation. *International Journal of Human Resource Management*, 17(3), 504-522.
- Kwan, V. S. Y., John, O. P., Robins, R. W., & Kuang, L. L. (2008). Conceptualizing and assessing self-enhancement bias: A componential approach. *Journal of Personality and Social Psychology*, 94(6), 1062-1077.
- LaHuis, D. M., & Avis, J. M. (2007). Using multilevel random coefficient modeling to investigate rater effects in performance ratings. *Organizational Research Methods*, 10(1), 97-107.

- Lam, S. S. K., & Schaubroeck, J. (1999). Total quality management and performance appraisal: An experimental study of process versus results and group versus individual approaches. *Journal of Organizational Behavior*, 20(4), 445-457.
- Lanza, P. (1985). Team appraisals. *Personnel Journal*, 64(3), 46-51.
- Latham, G. P., Budworth, M. H., Yanar, B., & Whyte, G. (2008). The influence of a manager's own performance appraisal on the evaluation of others. *International Journal of Selection and Assessment*, 16(3), 220-228.
- Latham, G. P., & Seijts, G. H. (1997). The effect of appraisal instrument on managerial perceptions of fairness and satisfaction with appraisals from their peers. *Canadian Journal of Behavioral Science*, 29(4), 275-282.
- Levine, R. E., Kelly, P. A., Karakoc, T., Haidet, P. (2007). Peer evaluation in a clinical clerkship: Students' attitudes, experiences, and correlations with traditional assessments. *Academic Psychiatry*, 31(1), 19-24.
- Levy, P. E., & Steelman, L. A. (1997). Performance appraisal for team-based organizations: A prototypical multiple rater system. In M. Beyerlein, D. Johnson, and S. Beyerlein (Eds.), *Advances in interdisciplinary studies of work teams: Team Implementation Issues, Vol.4* (pp.141-165). JAI Press, Greenwich, CT.
- Levy, P. E., & Williams, J. R. (2004). The social context of performance appraisal: A review and framework for the future. *Journal of Management*, 30(6), 881-905.
- Lievens, F., Conway, J. M., & De Corte, W. (2008). The relative importance of task, citizenship and counterproductive performance to job performance ratings: Do rater source and team-based culture matter? *Journal of Occupational and Organizational Psychology*, 81, 11-27.
- Lin, C., Standing, C., & Liu, Y. C. (2008). A model to develop effective virtual teams. *Decision Support Systems*, 45(4), 1031-1045.
- London, M. (2007). Performance appraisal for groups: Models and methods for assessing group processes and outcomes for development and evaluation. *Consulting Psychology Journal: Practice and Research*, 59(3), 175-188.
- London, M., Mone, E. M., & Scott, J. C. (2004). Performance management and assessment: Methods for improved rater accuracy and employee goal setting. *Human Resource Management*, 43(4), 319-336.

- Long, W. S., Long, E. J., & Dobbins, G. H. (1998). Correlates of satisfaction with a peer evaluation system: Investigation of performance levels and individual differences. *Journal of Business and Psychology*, 12(3), 299-312.
- Loughry, M. L., Ohland, M. W., & Moore, D. D. (2007). Development of a theory-based assessment of team member effectiveness. *Educational and Psychological Measurement*, 67(3), 505-524.
- Manshor, A. T., & Kamalanabhan, T. J. (2000). An examination of raters' and ratees' preferences in process and feedback in performance appraisal. *Psychological Reports*, 86(1), 203-214.
- Martell, R. F., & Evans, D. P. (2005). Source-monitoring training: Toward reducing rater expectancy effects in behavioral measurement. *Journal of Applied Psychology*, 90(5), 956-963.
- Martell, R. F., Guzzo, R. A., & Willis, C. E. (1995). A methodological and substantive note on the performance-cue effect in ratings of work-group behavior. *Journal of Applied Psychology*, 80(1), 191-195.
- Martins, L. L., Gilson, L. L., & Maynard, M. T. (2004). Virtual teams: What do we know and where do we go from here? *Journal of Management*, 30(6), 805-835.
- Marx, L. M., & Squintani, F. (2009). Individual accountability in teams. *Journal of Economic Behavior & Organization*, 72(1), 260-273.
- Maurer, T. J., Palmer, J. K., & Lisnov, S. S. (1995). Distinguishing context effects from context errors in judgments of behaviors. *Journal of Applied Social Psychology*, 25(18), 1637-1651.
- Maurer, T. J., Raju, N. S., & Collins, W. C. (1998). Peer and subordinate performance appraisal measurement equivalence. *Journal of Applied Psychology*, 83(5), 693-702.
- May, A., & Carter, C. (2001). A case study of virtual team working in the European automotive industry. *Int'l Journal of Industrial Ergonomics*, 27(3), 171-186.
- May, G. L. (2006). The effect of social style on peer evaluation ratings in project teams. *Journal of Business Communication*, 43(1), 4-20.
- Mero, N. P., Guidice, R. M., & Brownlee, A. L. (2007). Accountability in a performance appraisal context: The effect of audience and form of accounting on rater response and behavior. *Journal of Management*, 33(2), 223-252.

- Miller, D. L. (2001). Reexamining teamwork KSAs and team performance. *Small Group Research*, 32(6), 745-766.
- Miller, J. S. (2001). Self-monitoring and performance appraisal satisfaction: An exploratory field study. *Human Resource Management*, 40(4), 321-332.
- Miller, J. S., & Cardy, R. L. (2000). Self-monitoring and performance appraisal: Rating outcomes in project teams. *Journal of Organizational Behavior*, 21, 609-626.
- Moers, F. (2005). Discretion and bias in performance evaluation: The impact of diversity and subjectivity. *Accounting Organizations and Society*, 30(1), 67-80.
- Moore, D. A. (2007). When good = better than average. *Judgment and Decision Making Journal*, 2(5), 277-291.
- Morahan, M. J. (1996). Should peers' evaluations be used in class projects? Questions regarding reliability, leniency, and acceptance. *Psychological Reports*, 78(3), 1243-1250.
- Morgan, P. J., Pittini, R., Regehr, G., Marrs, C., & Haley, M. F. (2007). Evaluating teamwork in a simulated obstetric environment. *Anesthesiology*, 106(5), 907-915.
- Murphy, K. R., & DeShon, R. (2000). Interrater correlations do not estimate the reliability of job performance ratings. *Personnel Psychology*, 53(4), 873-900.
- Napoles, J. (2008). Relationships among instructor, peer, and self-evaluations of undergraduate music education majors' micro-teaching experiences. *Journal of Research In Music Education*, 56(1), 82-91.
- Netemeyer, R. G., & Maxham, J. G. (2007). Employee versus supervisor ratings of performance in the retail customer service sector: Differences in predictive validity for customers outcomes. *Journal of Retailing*, 83(1), 131-145.
- Norman, C. A., & Zawacki, R. A. (1991). Team appraisals: Team approach. *Personnel Journal*, 70(9), 101-103.
- Novicevic, M. M., Buckley, M.R., Harvey, M. G., & Fung, H. (2008). Self-evaluation bias of social comparisons in ethical decision making: The impact of accountability. *Journal of Applied Social Psychology*, 38(4), 1061-1091.
- Ostroff, C., Atwater, L. E., & Feinberg, B. J. (2004). Understanding self-other agreement: A look at rater and ratee characteristics, context, and outcomes. *Personnel Psychology*, 57(2), 333-375.

- Ozogul, G., Olina, Z., & Sullivan, H. (2008). Teacher, self and peer evaluation of lesson plans written by preservice teachers. *Educational Technology Research and Development*, 56(2), 181-201.
- Ozogul, G., & Sullivan, H. (2009). Student performance and attitudes under formative evaluation by teacher, self, and peer evaluators. *Educational Technology Research and Development*, 57(3), 393-410.
- Palmer, J. K., & Loveland, J. M. (2008). The influence of group discussion on performance judgments: Rating accuracy, contrast effects, and halo. *Journal of Psychology*, 142(2), 117-130.
- Patiar, A., & Mia, L. (2008). The effect of subordinates' gender on the difference between self-ratings, and superiors' ratings, of subordinates' performance in hotels. *International Journal of Hospitality Management*, 27(1), 53-64.
- Payne, S. C., Horner, M. T., Boswell, W. R., Schroeder, A. N., Stine-Cheyne, K. J., & Kelleen, J. (2009). Comparison of online and traditional performance appraisal systems. *Journal of Managerial Psychology*, 24(6), 526-544.
- Pimmel, R. L. (2003). A practical approach for converting group assignments into team projects. *IEEE Transactions on Education*, 46(2), 273-282.
- Price, K. H., Harrison, D. A., & Gavin, J. H. (2006). Withholding inputs in team contexts: Member composition, interaction processes, evaluation structure, and social loafing. *Journal of Applied Psychology*, 91(6), 1375-1384.
- Reb, J., & Greguras, G. J. (2010). Understanding performance ratings: Dynamic performance, attributions, and rating purpose. *Journal of Applied Psychology*, 95(1) 213-220.
- Roberson, L., Galvin, B. M., & Charles, A. C. (2007). When group identities matter in performance appraisal. *Academy of Management Annals*, 1, 617-650.
- Roberts, G. E. (1996). A case study in performance appraisal system development: Lessons from a municipal police department. *American Review of Public Administration*, 26(3), 361-385.
- Roberts, G. E. (2002). Employee performance appraisal system participation: A technique that works. *Public Personnel Management*, 31(3), 333-342.
- Roberts, G. E. (2003). Employee performance appraisal system participation: A technique that works. *Public Personnel Management*, 32(1), 89-98.

- Robins, R. W., & Beer, J. S. (2001). Positive illusions about the self: Short-term benefits and long-term costs. *Journal of Personality and Social Psychology*, 80(2), 340-352.
- Robbins, T. L., & Denisi, A. S. (1993). Moderators of sex bias in the performance-appraisal process- A cognitive analysis. *Journal of Management*, 19(1), 113-126.
- Roch, S. G. (2005). An investigation of motivational factors influencing performance ratings: Rating audience and incentive. *Journal of Managerial Psychology*, 20, 695-711.
- Roch, S. G. (2006). Discussion and consensus in rater groups: Implications for behavioral and rating accuracy. *Human Performance*, 19(2), 91-115.
- Roch, S. G., Ayman, R., Newhouse N., & Harris, M. (2005). Effect of identifiability, rating audience, and conscientiousness on rating level. *International Journal of Selection and Assessment*, 13(1), 53-62.
- Roch, S. G., & McNail, L. A. (2007). An investigation of factors influencing accountability and performance ratings. *Journal of Psychology*, 141(5), 499-523.
- Roch, S. G., Paquin, A. R., & Littlejohn, T. W. (2009). Do Raters Agree More on Observable Items? *Human Performance*, 22(5), 391-409.
- Roch, S. G., Sternburgh, A. M., & Caputo, P. M. (2007). Absolute vs relative performance rating formats: Implications for fairness and organizational justice. *International Journal of Selection and Assessment*, 15(3), 302-316.
- Rothrock, L., Cohen, A., Yin, J., Thiruvengada, H., & Nahum-Shani, I. (2009). Analyses of team performance in a dynamic task environment. *Applied Ergonomics*, 40(4), 699-706.
- Saavedra, R., & Kwun, S. K. (1993). Peer evaluation in self-managing work groups. *Journal of Applied Psychology*, 78(3), 450-462.
- Salas, E., Cooke, N. J., & Rosen, M. A. (2008). On teams, teamwork, and team performance: Discoveries and developments. *Human Factors*, 50(3), 540-547.
- Salas, E., Sims, D. E., & Burke, C. S.. (2005). Is there a big five in teamwork? *Small Group Research*, 36, 555-599.
- Sargent, C. N. (2001). Team performance evaluation for company officers. *Advanced Rescue Technology*, 4(5), 49-55.



- Sargeant, J., Mann, K., van der Vleuten, C., & Metsemakers, J. (2008). "Directed" self-assessment: Practice and feedback within a social context. *Journal of Continuing Education in The Health Professions*, 28(1), 47-54.
- Saturno, P. J., Palmer, R. H., & Gascon, J. J. (1999). Physician attitudes, self-estimated performance and actual compliance with locally peer-defined quality evaluation criteria. *International Journal for Quality in Health Care*, 11(6), 487-496.
- Savelsbergh, C. M. J. H., van der Heijden, B. I. J. M., & Poell, R. F. (2009). The development and empirical validation of a multidimensional measurement instrument for team learning behaviors. *Small Group Research*, 40(5), 578-607.
- Schleicher, D. J., Bull, R. A., & Green, S. G. (2009). Rater reactions to forced distribution rating systems. *Journal of Management*, 35(4), 899-927.
- Schroeder, J. Q. (1998). Investigating the effects of subtle ratee characteristics in team performance appraisal: The influence of personality and cohesiveness. Unpublished Masters thesis, Old Dominion University, US -- Virginia. Retrieved March 17, 2010, from WorldCat.
- Scott, S. G., & Einstein, W. O. (2001). Strategic performance appraisal in team-based organizations: One size does not fit all. *Academy of Management Executive*, 15(2), 107-116.
- Sedikides, C., & Gregg, A. P. (2008). Self-enhancement: Food for thought. *Perspectives on Psychological Science*, 3(2), 102-116.
- Sedikides, C., Herbst, K. C., Hardin, D. P., & Dardis, G. J. (2002). Accountability as a deterrent to self-enhancement: The search for mechanisms. *Journal of Personality and Social Psychology*, 83(3), 592-605.
- Sedikides, C., Horton, R. S., & Gregg, A. P. (2007). The why's the limit: Curtailing self-enhancement with explanatory introspection. *Journal of Personality*, 75(4), 783-824.
- Shore, T. H., & Taschian, A. (2007). Effects of feedback accountability and self-rating information on employee appraisals: A replication and extension. *Psychological Reports*, 100(3), 1091-1100.
- Simonds Jr., J. G., & Bell, R. C. (1997). Creating individual and team performance measurements. *Employee Relations Today*, 23(4), 27-35.

- Sisco, H., & Leventhal, G. (2007). Effect of prior performance on subsequent performance evaluation by field independent-dependent raters. *Perceptual and Motor Skills*, 105(3), 852-861.
- Smith, B. D., & Switzer, F. S. (2009). The effects of accountability on leniency reduction in self ratings. Unpublished master's thesis, Clemson University, Clemson, South Carolina.
- Smith, D. R., DiTomaso, N., Farris, G. F., & Cordero, R. (2001). Favoritism, bias, and error in performance ratings of scientists and engineers: The effects of power, status, and numbers. *Sex Roles*, 45(5-6), 337-358.
- Smith, A. F. R., & Fortunato, V. J. (2008). Factors influencing employee intentions to provide honest upward feedback ratings. *Journal of Business and Psychology*, 22(3), 191-207.
- Smither, J. W., London, M., & Reilly, R. R. (2005). Does performance improve following multisource feedback? A theoretical model, meta-analysis, and review of empirical findings. *Personnel Psychology*, 58(1), 33-66.
- Spies, T. H., Mokkink, H. G. A., Robbe, P. F. D., Grol, R. P. T. M. (2004). Which data source in clinical performance assessment? A pilot study comparing self-recording with patient records and observation. *International Journal for Quality in Healthcare*, 16(1), 65-72.
- Stark, E., & Poppler, P. (2009). Leadership, performance evaluations, and all the usual suspects. *Personnel Review*, 38(3), 320-338.
- Stoffey, R. W., & Reilly, P. R. (1997). Training appraisees to participate in appraisal: Effects on appraisers and appraisees. *Journal of Business and Psychology*, 12(2), 219-239.
- Stubblebine, P. C. (2001). Perception and acceptance of evaluations by supervisors and peers. *Current Psychology*, 20(1), 85-94.
- Summer, H. C., & Bilgic, R. (2006). Use of nontraditional rating sources in performance appraisals. *Turk Psikoloji Dergisi*, 157), 25-43.
- Sundvik, L., & Lindeman, M. (1998). Acquaintanceship and the discrepancy between supervisor and self-assessments. *Journal of Social Behavior and Personality*, 13(1), 117-126.

- Sung, Y. T., Chang, K. E., Chiou, S. K., & Hou, H. T. (2005). The design and application of a web-based self- and peer-assessment system. *Computers & Education*, 45(2), 187-202.
- Sweet, M. (2009). Individual and peer calibration in team-based testing. Ph.D. dissertation, The University of Texas at Austin, United States -- Texas. Retrieved February 26, 2010, from Dissertations & Theses: Full Text. (Publication No. AAT 3341963).
- Taggar, S., & Brown, T. C. (2006). Interpersonal affect and peer rating bias in teams. *Small Group Research* 37(1), 86-111.
- Tasa, K., Taggar, S., & Seijts, G. H. (2007). The development of collective efficacy in teams: A multilevel and longitudinal perspective. *Journal of Applied Psychology*, 92(1), 17-27.
- Taylor, E. K., & Wherry, R. J. (1951). A study of leniency in two rating systems. *Personnel Psychology*, 4, 39-47.
- Thorsteinson, T. J., Breier, J., Atwell, A., Hamilton, C., & Privette, M. (2008). Anchoring effects on performance judgments. *Organizational Behavior and Human Decision Processes*, 107(1), 29-40.
- Tziner, A., & Murphy, K. R. (1999). Additional evidence of attitudinal influences in performance appraisal. *Journal of Business and Psychology*, 13(3), 407-419.
- Tziner, A., Murphy, K. R., & Cleveland, J. N. (2001). Relationships between attitudes toward organizations and performance appraisal systems and rating behavior. *International Journal of Selection and Assessment*, 9(3), 226-239.
- Tziner, A., Murphy, K. R., & Cleveland, J. N. (2002). Does conscientiousness moderate the relationship between attitudes and beliefs regarding performance appraisal and rating behavior? *International Journal of Selection and Assessment*, 10(3), 218-224.
- Tziner, A., Murphy, K. R., & Cleveland, J. N. (2005). Contextual and rater factors affecting rating behavior. *Group & Organizational Management*, 30(1), 89-98.
- Tziner, A., Murphy, K., Cleveland, J. N., Yavo, A., & Hayoon, E. (2008). A new old question: Do contextual factors relate to rating behavior: An investigation with peer evaluations. *International Journal of Selection and Assessment*, 16(1), 59-67.
- Valami, H. B. (2009). Group performance evaluation, an application of data envelopment analysis. *Journal of Computational and Applied Mathematics*, 230(2), 485-490.

- Valle, M., & Bozeman, D. P. (2002). Interrater agreement on employees' job performance: Review and directions. *Psychological Report, 90*(3), 975-985.
- Valle, M., & Davis, K. (1999). Teams and performance appraisal: Using metrics to increase reliability and validity. *Team Performance Management, 5*(8), 238-244.
- Van Iddenkinge, C. H., Putka, D. J., Raymark, P. H., & Eidson, C. E. (2005). Modeling error variance in job specification ratings: The influence of rater, job, and organization-level factors. *Journal of Applied Psychology, 90*(2), 323-334.
- Villanova, P., Bernardin, H. J., Dahmus, S. A., & Sims, R. L. (1993). Rater leniency and performance appraisal discomfort. *Educational and Psychological Measurement, 53*(3), 788-799.
- Viswesvaran, C., Schmidt, F. L., & Ones, D. S. (2005). Is there a general factor in ratings of job performance? A meta-analytic framework for disentangling substantive and error influences. *Journal of Applied Psychology, 90*(1), 108-131.
- Viswesvaran, C. Schmidt, F. L., & Ones, D. S. (2002). The moderating influence of job performance dimensions on convergence of supervisory and peer ratings of job performance: Unconfounding construct-level convergence and rating difficulty. *Journal of Applied Psychology, 87*(2), 345-354.
- Waldman, D. A. (1997). Predictors of employee preferences for multirater and group-based performance appraisal. *Group & Organizational Management, 22*(2), 264-287.
- Walker H., & Joines M. (2004). A guide to peer appraisal. *Nursing Management, 11*(1), 22-24.
- Wallace, D. M., & Hinsz, V. B. (2009). Group Members as Actors & Observers in Attributions of Responsibility for Group Performance. *Small Group Research, 40*(1), 52-71.
- Webber, S. S., Chen, G., Payne, S. C., Marsh, S. M., & Zaccaro, S. J. (2000). Enhancing team mental model measurement with performance appraisal practices. *Organizational Research Methods, 3*(4), 307-323.
- Weisband, S., & Atwater, L. (1999). Evaluating self and others in electronic and face-to-face groups. *Journal of Applied Psychology, 84*(4), 632-639.
- Whiting, S. W., Podsakoff, P. M., & Pierce, J. R. (2008). Effects of task performance, helping, voice, and organizational loyalty on performance appraisal ratings. *Journal of Applied Psychology, 93*(1), 125-139.

- Whitmore, J. N. (2005). Assessing performance in complex team environments. *Aviation Space and Environmental Medicine*, 76(7), C31-C33.
- Williams, B. C., He, B. B., Elger, D. F., & Schumacher, B. E. (2007). Peer evaluation as a motivator for improved team performance in Bio/Ag Engineering design classes. *International Journal of Engineering Education*, 23(4), 698-704.
- Wilson, B. L., Anderson, J., Peluso, C., & Priest, J. (2009). Student satisfaction and team development outcomes with preassigned learning communities. *Journal of Professional Nursing*, 25(1), 15-22.
- Woehr, D. J., & Roch, S. G. (1996). Context effects in performance evaluations: The impact of ratee sex and performance level on performance ratings and behavioral recall. *Organizational Behavior and Human Decision Processes*, 66, 31-41.
- Woehr, D. J., Sheehan, M. K., Bennett, W. (2005). Assessing measurement equivalence across rating sources: A multitrait-multirater approach. *Journal of Applied Psychology*, 90(3), 592-600.
- Wong, K. F. E., & Kwong, J. Y. Y. (2007). Effects of rater goals on rating patterns: Evidence from an experimental field study. *Journal of Applied Psychology*, 92(2), 577-585.
- Wooten, T. C., & Brown, B. K. (1998). Upward performance appraisals for audit teams: Professional development that is simple. *CPA Journal*, 68(11), 60-62.
- Yun, G. J., Donahue, L. M., Dudley, N. M., & McFarland, L. A. (2005). Rater personality, rating format, and social context: Implications for performance appraisal ratings. *International Journal of Selection and Assessment*, 13(2), 97-107.
- Zigon, J. (1994). Making performance appraisals work for teams. *Training*, 31(6), 58-64.
- Zigon, J. (1995). Team Performance Measurement: A process for creating team performance standards. *Compensation & Benefits Review*, 29(1), 38-47.
- Zhang B., & Ohland, M. W. (2009). How to assign individualized scores on a group project: An empirical evaluation. *Applied Measurement In Education*, 22(3), 290-308.